Humanities and Social Sciences Communications

Article in Press

https://doi.org/10.1057/s41599-026-06941-6

Cruise service quality improvement: a quality function deployment approach with online reviews by large language models

Received: 21 April 2025

Accepted: 2 March 2026

Cite this article as: Gai, T., Wu, J., Xing, Y. *et al.* Cruise service quality improvement: a quality function deployment approach with online reviews by large language models. *Humanit Soc Sci Commun* (2026). https://doi.org/10.1057/s41599-026-06941-6

Tiantian Gai, Jian Wu, Yumei Xing, Yujia Liu & Mingshuo Cao

Cruise service quality improvement: A quality function deployment approach with online reviews by large language models

Abstract: Over the past decade, although the global demand for cruise tourism has steadily increased, a decline in customer satisfaction has also emerged as a significant challenge. Quality function deployment (QFD) is an effective approach for transforming customer requirements into product or service quality characteristics, which helps to identify key customer needs and optimize service quality. Therefore, this study aims to apply the QFD model to provide practical recommendations for improving cruise service quality. To achieve this, the large language models (LLMs) combined with prompt engineering is first utilized to extract from online cruise reviews and conduct sentiment analysis. Then, the Kano model is applied to classify customer requirements and the weight balance coefficient is introduced to ensure a rational allocation of weights. Hence, a social network-based bilateral interaction consensus mechanism is developed to resolve opinion conflicts within the QFD decision making team, enabling consensus-driven decisions and deriving the final prioritization of quality characteristics. Finally, a real-world cruise case study is conducted to validate the proposed approach, supported by systematic analysis and discussion to highlight its advantages. Overall, this study establishes a QFD framework that integrates LLMs, Kano model and group consensus methods based on online cruise reviews, which provides a data-driven and adaptive solution for improving cruise service quality.

Key words: Cruise service, Quality function deployment, Large language models, Kano model, Social network group consensus

1 Introduction

Cruise tourism is an essential part of the global tourism industry, occupying a significant position in the leisure and tourism market (Sun et al., 2023). Compared to the traditional land-based tourism products, cruise tourism involves a greater number of elements and service components, making it more complex (Sun et al., 2021). In recent years, although the industry has continued to grow and attract more tourists, it has also faced the challenges of declining customer satisfaction (Qorib et al., 2023) , which may be attributed to factors such as inadequate facility maintenance and fluctuations in service quality. Therefore, enhancing service quality and meeting diverse customer needs have become key issues that require attention within the industry. The quality function deployment (QFD) model is an effective planning approach designed to translate customer requirements (CRs) into specific product or service quality characteristics (QCs) (Gai et al., 2024a; Liu et al., 2022). By applying the QFD model, companies can systematically analyze customer needs, optimize service processes, and enhance the overall customer experience. Therefore, employing the QFD model to explore improvements in cruise service quality presents a feasible approach for strengthening industry competitiveness and increasing customer satisfaction.

The first research issue of the QFD related to cruise service is how to accurately identify customer requirements (CRs) by online review platforms. Cruise review websites, such as CruiseCritic, contain a vast amount of information related to perceived value of items, which is valuable in exploring customer needs and preferences (Ji et al., 2025; Wang et al., 2024b). Some studies have attempted to extract CRs from online reviews, for example, Liu et al. (2022) and Gai et al. (2024a) identified CRs using word frequency-based methods from customers' online shopping comments. However, these research primarily relies on traditional text processing techniques, which often struggle to capture contextual meaning and implicit customer needs. As further demonstrated in the comparative analysis later in this study, traditional methods such as TF-IDF, LDA, and KeyBERT show clear limitations in either efficiency or accuracy, reinforcing the need to explore more advanced approaches. The rapid development of large language models (LLMs) offers a new approach for accurately and efficiently extracting customer needs from large-scale and unstructured review data (Dagdelen et al., 2024). Compared to the traditional methods, LLMs can better understand textual context, identify implicit needs, and process more complex language expressions (Zhang et al., 2024). Additionally, their capabilities in multitask learning and contextual association analysis enable more effective customer feedback analysis and key demand identification. Therefore, the first aim of this article is to use the large language model (LLM) combined with prompt engineering to extract CRs from online cruise reviews and conduct sentiment analysis.

The second research issue of the QFD related to cruise service is how to determine the CRs' weights for the final prioritization. Currently, some studies have adopt different weight allocation methods to assign weights to CRs, such as the DEMATEL (Chen et al., 2021) and CRITIC method (Wang et al., 2024a). The Kano model classifies CRs into different categories (Kano et al., 1984), allowing a more precise reflection of customer perceptions, which provides supports for weight allocation by distinguishing the impact of different requirements. Some studies have integrated the Kano model into the analysis of CRs, Wu et al. (2021) identified the attribute category of CRs adopting the Kano model. He et al. (2021) developed qualitative and quantitative analysis of CRs using Kano model to generate CR's weight. However, most of the existing methods rely on expert subjective evaluations, and they lack analysis based on objective data sources. Such reliance on subjective judgments may lead to less consistent or less generalizable results, as the outcomes can be influenced by the specific composition of the expert group and their individual perspectives. The second aim of this article is to conduct Kano classification based on objective data such as online reviews to optimize weight allocation, which can reduce subjective bias, improve the reliability of demand classification, and make weight allocation more dynamic and accurate.

The third research issue of the QFD related to cruise service is how to help group experts to reach consensus. Typically, experts from different departments or backgrounds form a decision making group to collaboratively complete the QFD process, which can be regarded as a group decision making (GDM) problem (Wang et al., 2024a). GDM leverages the collective wisdom of experts (Cao et al., 2025; Yu et al., 2025), leading to more reliable QFD results compared to individual decision making. However, many existing studies overlook the essence of GDM within the QFD model and fail to adequately consider the social network relationships among decision experts, they still rely on direct aggregation methods, such as simple averaging or fixed weighting, which may lead to less consistent or less acceptable outcomes. For example, Fang et al. (2020) developed a QFD framework based on rough cloud model theory. Wu at al. (2021) using extended TOPSIS method to derive the final QFD results. Therefore, the incorporation of structured consensus mechanisms and trust relationships in QFD studies remains relatively limited. Due to differences in their positions and experiences, the evaluation information provided by experts may be inconsistent, and directly aggregating these opinions could result in unacceptable or unreasonable outcomes. The consensus reaching process (CRP) in GDM is an effective tool of eliminating conflicting opinions and obtaining a consensus-based decision result (Zhou et al., 2024). Additionally, the social trust relationships between experts are essential for fostering cooperation among them, which facilitate the achievement of group consensus (Gai et al., 2024b; Xu et al., 2024). The third aim of this article is to propose a new interaction mechanism with social network to enhance the consensus achievement among QFD decision experts.

Based on the above discussion, it can be concluded that it is a promising approach to apply the QFD model to provide guidance for cruise service quality improvement. However, the existing QFD research still has certain limitations that need to be addressed: 1) Limitations in CRs extraction methods. Existing studies struggle to accurately capture contextual meaning and implicit needs, leading to less precise and comprehensive extraction results. 2) Lack of objective data support in CRs weight allocation. Although some studies incorporate the Kano model for CRs classification and weight allocation, they still lack large-scale data-driven quantitative analysis, which affects the reliability of classification results and the rationality of weight distribution. 3) Insufficient consideration of group consensus and social network relationships of decision experts. The absence of an effective consensus mechanism may lead to the direct aggregation of conflicting expert opinions, which can undermine the rationality and acceptability of the final decision. These three research issues are not isolated but are successive stages within the QFD methodological framework. The identification of CRs provides the fundamental input, weight allocation determines their relative importance in shaping service priorities, and the consensus mechanism ensures that these results are collectively acceptable to the expert group. Theses steps form a coherent process of a unified QFD methodological framework.

To address these challenges, this article aims to integrate LLMs, Kano model and group consensus method to construct a QFD model based on online reviews of cruise ships to provide reference guidance for improving cruise service quality. The contributions of the article are summarized as follows:

(1) This study leverages LLMs combined with prompt engineering to extract CRs from online cruise reviews and conduct sentiment analysis, which can enhance the accuracy of implicit need identification and provides a more precise sentiment assessment, ensuring a more reliable data foundation for subsequent analysis.

(2) It classifies CRs using the Kano model, and introduces the weight balance coefficients for each Kano category to assign appropriate weights for CRs based on objective sentiment analysis data. This approach enables a more objective and precise reflection of CRs' importance, ensuring a scientifically sound and rational weight distribution, thereby enhancing the reliability of QFD analysis.

(3) It proposes a bilateral interaction mechanism with social network to enhance the consensus achievement among QFD decision experts. By constructing an expert trust network and refining the opinion aggregation mechanism, this model improves the level of consensus in QFD decision making, ultimately leading to more acceptable and stable QFD decision outcomes.

The remainder of the paper is organized as follows: Section 2 provides a comprehensive review of the relevant literature. Section 3 provides preliminaries about Kano model, QFD model and social trust network. Section 4 introduces the methodological framework of the proposed QFD model for cruise service quality improvement. In Section 5, a case study of cruise ship Norwegian Breakaway is provided to illustrate the applicability of the proposed method, and Section 6 presents some analysis and discussion. Finally, Section 7 concludes this paper and suggests potential future research directions.

2 Literature review

This section provides a comprehensive review and analysis of relevant literature in three research fields: (1) research on cruise online reviews, (2) text processing of online reviews, and (3) consensus mechanism in social network group decision making (SNGDM), which form the theoretical basis of this research.

2.1 Research on cruise online reviews

Over the past decade, the demand for cruise tourism has continued to increase worldwide, and the cruise industry has become one of the most dynamic industries in the global economy (Chrysafis et al., 2024). As more tourists embark on cruise vacations, an increasing number of online platforms have emerged, websites such as CruiseCritic allow travelers to share their experiences, express their satisfaction levels, and voice their expectations. These online reviews generate a vast amount of user-generated content (UGC), and they have become a reliable resource for researchers studying cruise service quality and customer needs, and is crucial for cruise enterprise management decisions (Ji et al., 2025).

Some researchers have utilized cruise online reviews to analyze customer sentiment and satisfaction (Ahn et al., 2023). Sun et al. (2023) utilized sentiment analysis on over 97,000 online reviews from 172 cruise ships across 16 brands to examine how tourists express their sentiments about various aspects of the cruise experience and the differences in market positioning. Ma et al. (2024) proposed a deep learning-based cruise satisfaction evaluation model that integrates online cruise reviews with a large group consensus model, leveraging sentence-level information to comprehensively capture the attributes that users genuinely care about. In addition, some scholars have focused on exploring customer loyalty behavior and perceptions of value for money based on review data (Tiutiu et al., 2025). Castillo-Manzano et al. (2022) applied input-output circular plots and discrete choice models to identify the drivers that explain cruise tourism loyalty behavior based on over 150,000 online reviews posted by cruisers using electronic Word of Mouth (e-WOM). And Jiao et al. (2024) employed content analysis, regression analysis, and multi-group comparison on nearly 100,000 tourist reviews to examine the impact of cruisers' nine-dimensional experience on value-for-money perception and the differences across various market positions.

However, the rapid increase in cruise online reviews has led to information overload, making it challenging to extract relevant and valuable insights. To this end, Ji et al. (2025) developed an inherent-personalized review helpfulness evaluation model, which utilizes deep learning techniques to identify high-quality and highly relevant comments to support group cruise decision making. Additionally, while these reviews contain useful feedback, the process of systematically transforming textual data into actionable quality improvement measures remains underexplored. Existing studies primarily focus on analyzing review data to assess customer preferences and sentiments, but there is a lack of research on effectively utilizing this information to drive concrete service enhancements. This study addresses this gap by proposing a structured approach that integrates Kano and group consensus QFD model to convert cruise online review data into corresponding strategies for cruise service quality improvement.

2.2 Text processing of online reviews

With the rapid growth of social media, the volume and diversity of online reviews covering diverse products and services have significantly increased. Online reviews are typically presented as unstructured textual data, necessitating their transformation into structured information for a deeper understanding of consumer preferences, market trends, and strategic business decisions (Qorib et al., 2023). To address this, scholars have proposed various approaches for processing such data, which generally involve two key steps: topic extraction and sentiment analysis. Topic extraction identifies recurring themes within reviews, while sentiment analysis determines the emotional tendencies expressed by users. These approaches can be broadly classified into four categories: rule-based methods (Wu et al., 2022), statistical methods (Venugopalan et al., 2022), machine learning-based methods (Suresh Kumar et al., 2024), and deep learning-based methods (Qorib et al., 2023), which is listed in Table 1.

Table 1 Key methods for topic extraction and sentiment analysis.

Method category	Core concept	Classic methods
Rule-based methods	Utilize predefined linguistic rules, lexicons, or syntactic structures to identify topics or sentiments	Pattern matching, dependency parsing, sentiment lexicons
Statistical methods	Leverage frequency-based and co-occurrence patterns to infer topics or sentiments	TF-IDF, LDA, TextRank
Machine leaning-based methods	Treat topic extraction and sentiment analysis as supervised or unsupervised classification tasks	SVM, Random Forest, K-means clustering
Deep leaning-based methods	Use neural networks for end-to-end modeling, capturing semantic and contextual relationships	LSTM, BERT, GPT

Rule-based and statistical methods represent the earliest approaches to processing online reviews. Rule-based methods, such as sentiment lexicons or dependency parsing, are simple, interpretable, and computationally efficient. However, they are highly dependent on predefined dictionaries and linguistic rules, making them inflexible and less effective when dealing with the diverse and dynamic language used in customer reviews. Statistical methods, such as TF-IDF and LDA, improve upon this by capturing word frequency and co-occurrence patterns, allowing for more scalable topic extraction. Nevertheless, these methods primarily focus on surface-level text features and struggle to capture deeper contextual meanings or implicit customer needs. Compared with these approaches, LLMs provide a significant advancement, as they can leverage pre-trained knowledge to infer semantics beyond frequency patterns or predefined rules, thereby achieving stronger generalization and contextual understanding in extracting customer requirements.

In recent years, methods based on machine learning and deep learning have become the mainstream of online text processing technology. Ji et al. (2023) trained word vectors through Word2Vec to explore the distance between attribute dimensions and cluster centers to extract and verify user dements from online comments. Wan et al. (2024) proposed an Emotion-Cognitive Reasoning integrated BERT model (ECR-BERT), which enhances BERT's accuracy and explainability in sentiment analysis of online public opinions on emergencies by incorporating an emotion model and a self-adaptive fusion algorithm. Ma et al. (2024) introduced a deep learning-based cruise satisfaction evaluation model that combines online reviews with a large-group consensus approach, utilizing sentence-level attribute extraction and the pre-trained BART model for sentiment analysis.

In essence, machine learning relies on feature engineering, which is effective for text analysis but has limited capacity to capture contextual subtleties. Deep learning methods can automatically learn semantic relationships in text, enabling them to capture complex sentiment expressions and topic characteristics. However, they still face challenges such as dependence on large labeled datasets, high computational costs, and limited interpretability. LLMs like GPT-4, based on pre-training and fine-tuning techniques, can perform few-shot or zero-shot learning, enabling automatic reasoning and contextual understanding . This leads to more efficient topic extraction and sentiment analysis. Therefore, in this paper, LLM is employed to analyze cruise review data by prompt engineering, enabling more efficient extraction of in-depth insights embedded within the review data.

2.3 Consensus mechanism in SNGDM

The consensus mechanism in group decision making refers to the process by which decision makers refine their opinions through information interaction, ultimately converging on a unified decision outcome. Establishing group consensus is essential for enhancing the reliability of the decision outcome and ensuring its acceptability among decision participants (Gai et al., 2024b).

Traditional consensus interaction mechanisms primarily focus on exploring the specific rules or paradigms that guide how individuals or subgroups continuously engage in discussions to converge towards the group. Zha et al. (2019) pointed out that two consensus rules are commonly used in in group consensus decision making: (1) the identification rule (IR) and the direction rule (DR) and (2) the optimization based consensus rules. Guo et al. (2024) extended the traditional consensus models into multi-dimensional and multi-round minimum cost consensus model and presented and corresponding algorithm and consensus reaching process.

With the rapid development of social networks, complex and diverse social relationships have emerged among decision makers, a research focus has become how to enhance interaction willingness and foster group consensus by establishing and utilizing the social relationships among decision makers. Liu et al. (2023) argued that social trust network (STN) is the basis of interaction for decision makers, and proposed a STN-based multi-attribute group decision making consensus decision framework. Zhou et al. (2024) explored trust network-based group decision making, with a focus on how to use ordinal consensus measures to assess and facilitate the achievement of consensus. Dong et al. (2024) originally proposed the social network DeGroot (SNDG) model, establishing a great connection between opinion dynamics, social networks, and group decision making.

Overall, the aforementioned studies explored the role and mechanisms of factors such as trust relationships in facilitating group consensus. However, they primarily focus on the process of individual convergence toward group consensus and do not address how to effectively manage disagreements between individuals. To address this gap, the bilateral interaction consensus mechanism is proposed to identify and address key conflicts between decision makers, thereby enabling more efficient consensus achievement (Gai et al., 2024a). This paper develops a QFD model that integrates a bilateral interaction consensus mechanism in SNGDM, which can efficiently resolves conflicting opinions among QFD decision makers, and generate consensus-based and reliable QFD model outputs.

3 Preliminaries

In order to make this article self-contained, this section provides the preliminary knowledge related to the Kano model, the QFD model, and the social trust network.

3.1 Kano model

The Kano Model, developed by Kano et al. (1984), is a product management tool designed to analyze and categorize CRs. By evaluating CRs across two key dimensions: the degree of fulfillment and customer satisfaction, the model classifies CRs into different categories, which is shown in Figure 1. The five Kano categories are introduced as follows:

(1) Must-be quality (M). These are the fundamental requirements of customers. Their absence causes strong dissatisfaction, but their presence does not increase satisfaction.

(2) Attractive quality (A). These are unexpected features that customers do not anticipate. While their absence does not cause dissatisfaction, their presence significantly enhances satisfaction.

(3) One-dimensional quality (O). The level of customer satisfaction is positively correlated with their fulfillment level. The higher the level of fulfillment, the greater the customer satisfaction, and vice versa.

(4) Reversal quality (R). These are features that customers are not satisfied with. Providing them may actually decrease customer satisfaction.

(5) Indifferent quality (I). These features have no impact on customer satisfaction, regardless of whether they are present or not.

Existing studies have mainly studied the Kano model through standardized questionnaires to analyze the Kano categories of requirements, providing companies with deeper insights into the key drivers of customer satisfaction.

Figure 1 Kano model.

3.2 QFD model

QFD is a customer-oriented quality management approach widely applied in industries such as manufacturing, software development, and services (Liu et al., 2022). It focuses on systematically translating CRs into QCs, where QCs represent the essential attributes of product and service quality. In the QFD process, determining the prioritization of QRs is a critical decision making issue (Chen et al., 2021).

In classic QFD, establishing a house of quality (HOQ) to achieve the transformation from CRs to QCs is a fundamental and strategic method (Wu et al., 2021). The classical structure of a HOQ consists of six critical elements: CRs, importance of CRs, QCs, correlations among QCs, relationships between CRs and QCs, and the final priorities of QCs, as depicted in Figure 2.

Figure 2 House of quality.

3.3 Social trust network

Social network analysis (SNA) is an effective methodology for investigating the relationships between different social entities (Wasserman, 1994). Generally, trust relationships are considered a reliable form of social network relation and are widely applied in the research of SNGDM.

Let $E = \{e_1, e_2, \ldots, e_m\}$ denote a set of m decision experts, a social trust network can be modeled with a graph $G = \{E, L\}$ of nodes E and edges L, in which $L = \{(l_{h\to k}) | e_h, e_k \in E; h \neq k\}$ represents the set of trust relationships between decision experts (Bondy and Murty, 1976). In real-world scenarios, the expression of trust relationships between individuals often involves uncertainty, and the trust function is an effective method for modeling trust with uncertainty, and it is defined as follows.

Definition 1. (Victor et al. 2009) A trust function (TF) is defined as a tuple of the type $\lambda_{h\to k} = \{t_{h\to k}, d_{h\to k}\} \in [0,1]^2$, with '$t_{h\to k}$' and '$d_{h\to k}$' representing the trust degree and distrust degree from e_h to e_k.

Definition 2. (Victor et al. 2009) The trust score (TS) associated to a trust function $\lambda_{h\to k} = \{t_{h\to k}, d_{h\to k}\}$ is defined as follows

$$TS(\lambda_{h\to k}) = \frac{t_{h\to k} - d_{h\to k} + 1}{2}. \tag{1}$$

4 QFD model for improving cruise service quality

This section introduces the methodological framework of the proposed QFD model for cruise service quality improvement, which integrates theories and methods of LLMs, Kano model and social network group consensus decision making. Figure 3 depicts the framework of the proposed approach, consisting of the following three stages:

Stage A: Cruise CRs extraction and sentiment analysis based on LLMs.

Stage B: Classification and weighting of cruise CRs based on Kano model.

Stage C: QFD model with bilateral interaction consensus mechanism.

Figure 3 Framework of the proposed QFD model for cruise service quality improvement.

4.1 Text processing of cruise online reviews based on LLMs

This study first employs web crawler technology to collect the latest cruise review data concerning the target cruise ship from CruiseCritic, a platform that aggregates a vast number of user-generated reviews and detailed feedback on various cruise experiences. Usually, the unstructured text reviews need to be converted into a structured format, and the LLM is utilized in this paper to process the reviews. Compared to traditional machine learning methods, LLMs possess strong natural language understanding capabilities, enabling efficient processing of large-scale text data without requiring extensive manual annotation, thus enhancing both efficiency and accuracy. The objective of this section is to identify the key CRs and perform a detailed analysis of the sentiment polarity associated with each requirement utilizing prompt engineering for LLMs, providing a more reliable data foundation for subsequent analysis.

4.1.1 Prompt engineering for LLMs

Prompt engineering refers to the process of designing and optimizing input prompts for LLMs to guide them in generating high-quality and accurate responses (Zhang et al., 2024). A well-structured prompt can enhance task performance, reduce bias, and improve the controllability of the model. The key steps for using prompt engineering to guide LLM output can be summarized as follows:

Step 1. Task defining and role assignment. Explicitly describe the tasks to be completed and assign a specific identity to the model to enhance the professionalism and consistency of the output.

Step 2. Few-shot learning. Including examples or illustrations in the prompt helps the model understand the task requirements and improve accuracy.

Step 3. Specifying output format. Defining a structured output format, such as JSON or tables, facilitates better interpretability and downstream processing.

Step 4. Instruction Optimization. Refining the instructions to enhance precision, ensure logical coherence, and minimize errors in model outputs.

By following these structured steps, prompt engineering can significantly improve the accuracy and efficiency of data processing, enhancing the practical utility of LLMs in real-world applications.

4.1.2 Cruise CRs extraction and sentiment orientation analysis

In this section, we employ the LLM ChatGPT-4o to extract attributes and perform sentiment analysis on the collected cruise online review data. The model was accessed via the OpenAI Python client (Python 3.8.10, openai==1.37.0). To ensure deterministic outputs, all runs were executed with temperature set to 1, top_p set to 1, and max_tokens limited to 2,000, the overview of task prompting is described in Figure 4, and the complete prompt is provided in Appendix A (see Supplementary Files). Specifically, to identify critical CRs, the LLM' role is first defined as a professional data analyst specializing in extracting core needs from customer feedback, and a corresponding prompt is designed to specify the task requirements. The prompt instructs the model to analyze the provided cruise customer reviews and extract top I key CRs, such as dining experience, facilities, etc. Additionally, for each identified CRs, the model is required to provide a concise explanation, and a structured output format for the extracted results is specified. After continuous optimization of the prompt, the final set of extracted CRs is represented as $CR = \{CR_1, CR_2, \ldots, CR_I\}$.

Building on the identification of key CRs, this study further utilizes LLM ChatGPT-4o to analyze the sentiment orientation of the cruise reviews. First, the model is assigned the role of a professional data analyst specializing in customer feedback analysis, focusing on sentiment classification for key requirements related to cruise experiences. Then the model is instructed to evaluate the provided cruise review data and determine the sentiment orientation corresponding to the predefined CRs for each review. Finally, to ensure consistency and facilitate further analysis, the model is required to output the results using a standardized set of sentiment labels, i.e., $SL = \{s_1 = `Very\ Negative', s_2 = `Negative', s_3 = `Neutral', s_4 = `Positive', s_5 = `Very\ Positive', N = `Not\ Mentioned\ or\ Not\ Evaluated'\}$.

The above systematic text processing approach based on LLMs enhances the accuracy and reliability of customer insight extraction and sentiment classification, facilitating further analysis and cruise service optimization.

Figure 4 Overview of prompting for cruise CRs extraction and sentiment orientation analysis.

4.2 Classification and weighting of cruise CRs based on Kano model

This section utilizes the Kano model to further analyze the results of CRs extraction and sentiment analysis obtained above, aiming to determine the specific classification of CRs, and further assign corresponding weights.

4.2.1 Kano classification of CRs

The construction of Kano model requires data that characterizes the degree of customer satisfaction and customer demand. The degree of customer demand for cruise quality attributes can be regarded as the level of attention paid to the specific CRs, which is associated with the frequency of mentions of the corresponding CRs in cruise reviews.

Suppose a total of z online cruise reviews have been collected, denoted as $R = \{r_1, r_2, \ldots, r_z\}$, from which the I most important CRs have been identified, denoted as $CR = \{CR_1, CR_2, \ldots, CR_I\}$. Let SL_k represent the sentiment label associated with CR_k, to quantify customers' attention to a specific CR, this study employs the following indicator to calculate the number of times CR_k is mentioned:

$$\vartheta_k = \begin{cases} 1, & if\ SL_k \neq N; \\ 0, & if\ SL_k = N. \end{cases} \tag{2}$$

And the customer attention (CA) to CR_k in $R = \{r_1, r_2, \ldots, r_z\}$ can be derived as

$$CA_k = \frac{\sum_{k=1}^{z} \vartheta_k}{z}. \tag{3}$$

Furthermore, based on the sentiment orientation analysis result of $R = \{r_1, r_2, \dots, r_z\}$ on each CR, an indicator can be constructed to represent customer satisfaction with each CR. Let $\chi_k = \sum_{k=1}^{z} \vartheta_k$, let $s_{\tau}^{k,h}(\tau \in \{1,2,\dots,g-1\}; h = 1,2,\dots,\chi_k)$ denote the sentiment labels associated with CR_k across all reviews $R = \{r_1, r_2, \dots, r_z\}$, the calculation of linguistic words is converted into the operation of subscripts, $\Delta^{-1}(s_\tau) = \tau$. Then the customer satisfaction (CS) to CR_k in $R = \{r_1, r_2, \dots, r_z\}$ can be obtained as

$$CS_k = \frac{\sum_{h=1}^{\chi_k} \Delta^{-1}\left(s_{\tau}^{k,h}\right)}{\chi_k \cdot g}. \tag{4}$$

It can be inferred that both CA_k and CS_k are defined as proportions or normalized averages, thus $CA_k, CS_k \in [0,1] (k = 1,2,\dots,I)$. Based on the definitions of different categories of requirements in the Kano model (as outlined in Section 3.1), CRs can be classified using a quadrant-based approach based on customer satisfaction and customer attention levels.

For the extracted $CR = \{CR_1, CR_2, \dots, CR_I\}$, the decision criteria for quadrant-based Kano classification are defined by the mean values of customer attention ($\overline{CA} = \frac{\sum_{k=1}^{I} CA_k}{I}$) and customer satisfaction ($\overline{CS} = \frac{\sum_{k=1}^{I} CS_k}{I}$). Specifically, $\overline{CA}$ is used as the threshold on the horizontal axis, and $\overline{CS}$ is used as the threshold on the vertical axis, since CRs with values above the average level are considered to attract relatively high attention or satisfaction, while those below the average level are regarded as receiving relatively low attention or satisfaction. This approach ensures that the classification is data-driven and adaptive to the characteristics of the specific dataset, avoiding predefined or subjective thresholds. In addition, the mean values reflect the overall central tendency of customer perceptions, providing a direct and interpretable benchmark for distinguishing relatively high and low attention or satisfaction levels within the dataset. Through this method, the must-be qualities, one-dimensional qualities, attractive qualities, and indifferent qualities can be assigned to distinct quadrants, as shown in Figure 5.

Figure 5 CR classification diagram based on Kano model.

Let ψ_k denote the Kano category to which CR_k belongs, the four quadrants is interpreted as follows:

(1) Must-be quality ($\psi_k = M$) is characterized by both high customer satisfaction and high customer attention. From the perspective of product reviews, must-be quality tend to receive considerable attention from users. Moreover, as essential features, they must align with users' psychological expectations, which is why their satisfaction levels are generally higher.

(2) One-dimensional quality ($\psi_k = O$) is characterized by low customer satisfaction but high attention levels. Customers expect this attribute to be better fulfilled, which explains the high level of attention. However, these attributes do not meet the psychological expectations of most users, leading to relatively low satisfaction. Optimizing these requirements can fulfill the majority of customer expectations and improve overall satisfaction.

(3) Attractive quality ($\psi_k = A$) is characterized by high customer satisfaction but low attention levels. These requirements evoke unexpected satisfaction in users, but they do not attract widespread attention

from most users, resulting in fewer mentions. Optimizing these attributes can address users' unexpected needs and further enhance their satisfaction.

(4) Indifferent quality ($\psi_k = I$) is characterized by both low customer satisfaction and low attention levels. Regardless of whether these attributes are optimized, customer satisfaction will remain unchanged.

4.2.2 Weight allocation of CRs

In the QFD model, the allocation of weights to CRs is a critical step, as it significantly influences the final ranking of QCs. In traditional QFD methods, the weighting of CRs is typically determined based on expert evaluation (Wu et al., 2021), which is inherently subjective. This study adopts a data-driven approach and integrates it with Kano model to objectively determine CRs' weights. After classifying CRs using Kano model, appropriate weights are assigned to different categories based on the classification results. In this way, the QFD model can more precisely reflect the significance of CRs, preventing the uniform treatment of all requirements and enhancing the accuracy of weight distribution.

In the Kano category, since indifferent attributes have no effect on improving customer satisfaction, CRs classified as indifferent attributes are directly eliminated in this paper and their weights are set to 0. Let w_k denote the weight of CR_k, $\forall k, \psi_k = I \Rightarrow w_k = 0$. Next, let CA_k and CS_k be as previously defined, the proposed weight index (WI) of CR_k is as follows:

$$WI_k = \alpha_k \cdot CA_k + \beta_k \cdot CS_k. \tag{5}$$

In which α_k and β_k serve as the weight balance coefficients between customer attention and customer satisfaction, $\alpha_k + \beta_k = 1$. Considering the characteristics of different Kano categories, the rules for setting weight balance coefficients α and β are as follows:

(1) For one-dimensional quality, a higher customer attention weight is assigned due to customers' elevated expectations, thus, $\alpha > \beta$ is set.

(2) For attractive quality, the impact is primarily reflected in satisfaction, warranting a higher customer satisfaction weight, thus, $\alpha < \beta$ is set.

(3) For must-be quality, attention and satisfaction weights are set equally to ensure that fundamental expectations are met, thus, $\alpha = \beta$ is set.

By introducing weight balance coefficients, the unique characteristics of one-dimensional and attractive qualities can be effectively reconciled, leading to a more rational allocation of CR weights. Finally, the weight index can be normalized to obtain the weight of each CR $w_k (k = 1,2, \dots, I)$:

$$w_k = \frac{WI_k}{\sum_{h=1}^{I} WI_h}. \tag{6}$$

4.3 QFD model with social network group consensus

In this section, the elements of the QFD model are first determined, followed by the construction of a bilateral interaction consensus model with social network. Finally, based on the consensus decision results obtained, the priority ranking of cruise service QCs is derived.

4.3.1 Determination of QFD model elements

This section focuses on defining the key elements of the QFD model with social network group consensus, including CRs, QCs, the evaluation matrix between CRs and QCs, decision experts, and the social trust relationships among decision experts.

First, building social network relationships among decision experts helps to explore their importance within the group and facilitates interaction of opinions and consensus building. Assuming that m cruise service-related departments $E=\{e_1,e_2,\dots,e_m\}$ constitute the QFD decision making group as decision experts. Experts use trust function $\lambda_{h\to k}=\{t_{h\to k},d_{h\to k}\}$ (defined in section 3.3) to express their trust in each other, forming the following social trust matrix (STM) among $E=\{e_1,e_2,\dots,e_m\}$:

$$STM=\begin{pmatrix} - & \lambda_{1\to 2} & \cdots & \lambda_{1\to m} \\ \lambda_{2\to 1} & - & \cdots & \lambda_{2\to m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m\to 1} & \lambda_{m\to 2} & \cdots & - \end{pmatrix} \tag{7}$$

Next, decision experts are required to generate J cruise service quality characteristics $QC=\{QC_1,QC_2,\dots,QC_J\}$ based on the I key customer requirements $CR=\{CR_1,CR_2,\dots,CR_I\}$ identified from online cruise review data.

Finally, decision experts are required to evaluate the relationship between CRs and cruise service QCs to construct the corresponding evaluation matrix. Let $f_{ij}^k\in[0,1]$ denote the evaluation value provided by the experts e_k on the relationship between CR_i and $QC_j(k=1,2,\dots,m;i=1,2,\dots,I;j=1,2,\dots,J)$, the evaluation matrix of expert e_k is represented as

$$F^k=(f_{ij}^k)_{I\times J}=\begin{pmatrix} f_{11}^k & f_{12}^k & \cdots & f_{1J}^k \\ f_{21}^k & f_{22}^k & \cdots & f_{2J}^k \\ \vdots & \vdots & \ddots & \vdots \\ f_{I1}^k & f_{I2}^k & \cdots & f_{IJ}^k \end{pmatrix} \tag{8}$$

4.3.2 Bilateral interaction consensus model in SNGDM

The CRP in GDM aims to reduce or eliminate disagreements among decision experts, ensuring the reliability and acceptability of the final decision results. In the QFD process, decision experts from various cruise service departments may provide diverse evaluations, leading to inconsistencies, so it is unreasonable to directly aggregate these evaluation opinions. To address this issue, this study employs a consensus mechanism of GDM with social network to resolve conflicts among QFD members, thereby achieving a consensus-based and reliable decision result.

Consensus mechanism serves as an effective approach to generate opinion modification references for experts to minimize discrepancies. Specifically, the bilateral interaction consensus mechanism is essential in identifying and resolving key conflicts between individuals, facilitating an efficient CRP within the group. The specific process of bilateral interaction consensus mechanism is introduced as follows.

Step 1. Consensus measure. First, it is necessary to calculate the consensus level of evaluation matrices between decision experts. Suppose the evaluation matrix of expert e_k and e_h are $F^k=(f_{ij}^k)_{I\times J}$ and $F^h=(f_{ij}^h)_{I\times J}$, the bilateral consensus level (BCL) between experts e_k and e_h is measured as

$$BCL_{kh}=1-\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\left|f_{ij}^k-f_{ij}^h\right|. \tag{9}$$

The general consensus level (GCL) of all experts $E=\{e_1,e_2,\dots,e_m\}$ is measured as:

$$GCL = \frac{2}{m(m-1)} \sum_{h=1,k>h}^{m-1} BCL_{kh}. \tag{10}$$

Equation (9) defines the BCL, which measures the consistency between the evaluation matrices of two experts e_k and e_h. It is calculated by assessing the differences in their evaluation matrices across all alternatives. A higher BCL value indicates stronger agreement between the two experts. Based on this, Equation (10) defines the GCL as the average BCL among all experts. The closer the GCL is to 1, the higher the overall agreement within the expert group.

Step 2. Consensus control. In order to determine whether the obtained group consensus level reaches the acceptable consensus, GCL is compared with a pre-defined consensus threshold $\gamma \in [0.5,1]$. If $GCL \geq \gamma$, then an acceptable level of consensus has been reached and the experts' evaluation matrices can be aggregated to derive the prioritization of cruise service QCs. Otherwise, the consensus mechanism needs to be implemented to generate modification suggestions and improve group consensus.

Step 3. Construction of consensus rules. The bilateral interaction consensus mechanism aims to eliminate key conflicts among experts through bidirectional negotiation and interaction, which has been proven effective in improving the overall rationality and acceptability of group decisions (Cao et al., 2024). Typically, the process starts by addressing the most significant disagreement between experts as it offers the greatest potential for enhancing consensus. Suppose e_p and e_q are identified with the greatest disagreement, $\{(e_p, e_q) | BCL_{pq} = \min_{h,k=1,\dots,m} BCL_{hk}\}$, let η_p and η_q be the interaction parameters used to control the acceptance of recommended information, then the recommended opinions for e_p and e_q, $\overline{F}^p = (\overline{f}^p_{ij})_{I \times J}$ and $\overline{F}^q = (\overline{f}^q_{ij})_{I \times J}$, can be obtained applying the following bilateral interaction consensus rules:

$$\begin{aligned} \overline{f}^p_{ij} &= \eta_p \cdot f^q_{ij} + (1-\eta_p) \cdot f^p_{ij}; \\ \overline{f}^q_{ij} &= \eta_q \cdot f^p_{ij} + (1-\eta_q) \cdot f^q_{ij}. \end{aligned} \tag{11}$$

Step 4. Establishment of consensus model. In the process of bidirectional interaction, experts usually do not fully accept the proposed modifications in order to maintain their initial positions. Therefore, it is necessary to establish an indicator to quantify the degree of collaboration among experts. In SNGDM, the level of consensus and trust between individuals are the core factors that affect their willingness to interact. Hence, the collaboration index (CI) between experts constructed in the article is as follows:

$$\begin{aligned} CI_{p\to q} &= \sqrt{BCL_{pq} * TS(\lambda_{p\to q})}; \\ CI_{q\to p} &= \sqrt{BCL_{pq} * TS(\lambda_{q\to p})}. \end{aligned} \tag{12}$$

Equation (12) defines the CI, which combines the bilateral consensus level (BCL) with the trust score (TS) between two experts. Intuitively, this index reflects both the similarity of their evaluation results and the degree of mutual trust . A higher CI value indicates that the two experts not only provide consistent assessments but also maintain stronger trust, making them more likely to collaborate effectively during the CRP.

Then the following bilateral interaction consensus model for maximum consensus between e_p and e_q within the constraints of their collaboration index is established:

$$max \;\; BCL^*_{pq} = 1 - \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| \bar{f}^p_{ij} - \bar{f}^q_{ij} \right| \tag{13}$$

$$\text{s.t.}\begin{cases}\bar{f}_{ij}^{p}=\eta_p\cdot f_{ij}^{q}+\left(1-\eta_p\right)\cdot f_{ij}^{p} & (13.1)\\ \bar{f}_{ij}^{q}=\eta_q\cdot f_{ij}^{p}+\left(1-\eta_q\right)\cdot f_{ij}^{q} & (13.2)\\ 0\le\eta_p\le\theta\cdot CI_{p\to q} & (13.3)\\ 0\le\eta_q\le\theta\cdot CI_{q\to p} & (13.4)\\ 0\le\theta\le 1,\eta_p+\eta_q\le 1 & (13.5)\end{cases}$$

In this model, the objective function aims to maximize the bilateral consensus level between two interacting experts. Equations (13.1–13.2) define the feedback rules, showing how each expert adjusts their evaluations by partially incorporating the other's opinion. Equations (13.3–13.4) constrain the range feedback parameters, with the upper bounds determined by the CI and scaled by $\theta\in[0,1]$, which reflect the level at which CI affects interaction parameters. Under the same CI, the larger θ, the higher the upper bound of interaction degree. Finally, the condition $\eta_p+\eta_q\le 1$ is introduced to prevent excessive adjustments from both sides, ensuring a balanced and stable consensus process.

Proposition 3.1. After implementing the bilateral feedback rules (11), the bilateral consensus level between the identified expert pair (e_p,e_q) increases, i.e.,

$$BCL_{pq}<BCL_{pq}^{*}\le 1.$$

Proof. From Eq. (9), we have

$$BCL_{pq}^{*}=1-\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\left|\bar{f}_{ij}^{p}-\bar{f}_{ij}^{q}\right|=1-\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\left|1-\eta_p-\eta_q\right|\left|f_{ij}^{p}-f_{ij}^{q}\right|$$

$$=1-\left|1-\eta_p-\eta_q\right|\cdot(1-BCL_{pq}).$$

Since $\eta_p,\eta_q>0,\eta_p+\eta_q\le 1$, then $0\le\left|1-\eta_p-\eta_q\right|\le 1$, therefore, $0\le\left|1-\eta_p-\eta_q\right|\cdot(1-BCL_{pq})\le 1-BCL_{pq})$, which means $BCL_{pq}<BCL_{pq}^{*}\le 1$.

This proposition shows that the bilateral interaction mechanism increases the consensus degree after each iteration, indicating that the proposed model progressively improves agreement and converges toward a higher consensus state.

4.3.3 Prioritization of cruise service QCs

Once the QFD decision experts reach a group consensus, the collective opinion based on consensus can be derived by aggregating the evaluation matrix of each department. To assign appropriate weights to decision experts, social network analysis is performed on their trust network. Let $TS(\lambda_{h\to k})$ denote the trust score of e_h on e_k, the in-centrality degree (ICD) of e_k is defined as

$$ICD_k=\frac{1}{m-1}\sum_{h=1,h\ne k}^{m}TS(\lambda_{h\to k}). \tag{14}$$

And the weight vector of decision experts $E=\{e_1,e_2,\ldots,e_m\}$ can be determined as

$$\pi_k = \frac{ICD_k}{\sum_{h=1}^{m} ICD_h}. \tag{15}$$

Then the final collective evaluation matrix on the relationship between CR_i and $QC_j(i = 1,2,\dots,I, j = 1,2,\dots,J)$ is obtained as

$$F^C = (f_{ij}^C)_{I\times J} = \sum_{k=1}^{m} \pi_k \overline{F}^k. \tag{16}$$

Let $w_i(i = 1,2,\dots,I)$ denote the weight of CRs obtained in section 4.2.2, the final score (FS) of cruise service $QC_j(j = 1,2,\dots,J)$ can be derived as

$$FS_j = \sum_{i=1}^{I} w_i f_{ij}^C. \tag{17}$$

The great the value of FS_j, the higher the priority of QC_j, and the final prioritization of QCs can be correspondingly obtained.

5 Case study

In this section, the proposed methodological framework is applied to the service quality improvement process of a real cruise ship to illustrate its practicality and effectiveness.

5.1 Research background

This study first collects cruise review data from CruiseCritic, a globally renowned platform for cruise reviews and discussions. CruiseCritic aggregates a substantial number of authentic reviews from passengers regarding various cruise ships, offering detailed evaluations of overall service quality. Owing to its broad coverage, large data volume, and rigorous verification mechanisms that ensure the authenticity and impartiality of the reviews, CruiseCritic serves as a highly authoritative and valuable resource in the cruise industry.

The Norwegian Breakaway is selected as the subject of analysis. Operated by Norwegian Cruise Line (NCL), the Norwegian Breakaway is a large cruise ship constructed by Germany's Meyer Werft shipyard. The Norwegian Breakaway represents a typical and mature case within the contemporary cruise industry. Its large operational scale and well-established management practices make it broadly representative of mainstream cruise operations. In addition, passenger reviews and service evaluation data for this ship are relatively abundant and complete, providing a reliable empirical basis for analysis.

As a modern luxury cruise vessel, the Norwegian Breakaway boasts numerous advantages, such as a wide array of entertainment facilities and diverse dining options. However, in recent years, the ship has also faced several challenges, including inconsistent service quality, inadequate maintenance of certain facilities, and fluctuations in passenger satisfaction.

Therefore, the objective of this paper is to extract the latest reviews concerning the Norwegian Breakaway from the CruiseCritic platform, identify the key customer requirements and perform sentiment analysis based on LLMs. Based on the obtained analysis results, the CRs are further classified and assigned with appropriate weights leveraging the Kano model. Additionally, a QFD model based on social network group consensus is established to derive the final ranking of cruise service QCs. Through this process, this research aim to provide scientific guidance and targeted recommendations for enhancing service quality within the cruise industry.

5.2 Model application and calculation

Following the specific procedures introduced in section 4, the application and calculation of the proposed QFD model on cruise ship Norwegian Breakaway is presented as follows.

Stage A. CRs extraction and sentiment analysis of cruise online reviews. On December 13, 2024, the latest 500 user reviews of Norwegian Breakaway were collected from the CruiseCritic platform using a web crawler. These reviews span from February 2019 to December 2024, representing recent passenger feedback over a five-year period. Duplicate entries were removed during preprocessing to ensure data consistency. The dataset consists only of publicly available user-generated content without personal identifiable information. Given that LLMs have been pre-trained on vast amounts of data, it is standard practice to divide the data into validation and test sets for effective output evaluation. To control potential bias, the dataset used in this study was independently collected and employed only for inference and validation, without any involvement in model training. The test set is used to assess the true performance of the LLM, and no adjustments can be made to the test data during this evaluation. The test data does not participate in prompt optimization and is used to evaluate the final generalization ability of the model. The validation set is manually annotated to serve as reference answers for the model's outputs. If the outputs of the model are acceptable, the effectiveness and practicality of the model are verified.

Specifically, the article first randomly selects 100 reviews $R = \{r_1, r_2, \ldots, r_{100}\}$ from the 500 collected data as the test set, and the determined prompts are applied to GPT-4o to perform CRs extraction and sentiment analysis. The outputs on the test set are presented in Tables 2 and 3. Next, an additional 100 reviews are randomly sampled from the remaining dataset to form the validation set, which is manually annotated by single annotator to serve as reference labels, as shown in Table 4. The validation results are used for comparative analysis and will be presented in the following section.

Table 2 Extracted CRs and descriptions.

CRs	Explanation	Aspects
Dining Experience ($\boldsymbol{CR_1}$)	Quality, variety, and service of both complimentary and specialty dining	Main dining rooms, buffets, specialty restaurants, and room service.
Cleanliness and Maintenance ($\boldsymbol{CR_2}$)	The condition of the ship and cleanliness of public areas and staterooms.	Overall cleanliness, maintenance of facilities, and hygiene standards.
Staff and Customer Service ($\boldsymbol{CR_3}$)	The interaction with crew and staff throughout the cruise.	Friendliness, helpfulness, efficiency, and professionalism of staff
Onboard Activities & Entertainment ($\boldsymbol{CR_4}$)	Entertainment options, shows, activities, and events on the ship.	Daytime activities, evening shows, poolside entertainment.
Embarkation & Disembarkation ($\boldsymbol{CR_5}$)	The process of boarding and leaving the ship.	Efficiency, organization, wait times, and communication at ports.
Excursions & Port Visits ($\boldsymbol{CR_6}$)	Activities and tours offered at destinations of the cruise visits.	Variety, organization, value, and timing of excursions.
Ship Layout & Facilities ($\boldsymbol{CR_7}$)	The design and functionality of the ship's amenities.	Pool areas, deck space, dining venue locations, and accessibility.
Cabin Experience ($\boldsymbol{CR_8}$)	The conditions and features of the staterooms.	Space, amenities, comfort, and noise levels.

Communication and Organization ($\boldsymbol{CR_9}$)	Effectiveness of communication from the cruise line before and during the cruise.	Information on schedules, changes, procedures, and ship events.
Pricing and Value for Money ($\boldsymbol{CR_{10}}$)	The perceived value of the cruise experience relative to its cost.	Costs of onboard services, excursions, and overall cruise fare.

Table 3 Sentiment analysis results on CRs.

	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6	CR_7	CR_8	CR_9	CR_{10}
r_1	s_3	N	s_4	s_2	s_1	s_2	s_1	N	s_4	s_2
r_2	s_5	N	s_4	s_4	s_2	s_2	N	N	s_2	s_1
...	...	...	...	...	...	...	...	...	...	...
r_{100}	s_2	s_2	s_1	N	s_1	N	s_2	s_1	s_1	s_1

Table 4 The structure of validation set.

Online cruise review	Template
1st time cruise. Waited in extreme long lines to tender off...	The sentiment orientation of CR_1 is s_2
1st time cruise. Waited in extreme long lines to tender off...	The sentiment orientation of CR_2 is N
…	…
1st time cruise. Waited in extreme long lines to tender off...	The sentiment orientation of CR_{10} is s_1

Stage B. Classification and weighting of CRs. Based on the obtained CR extraction and sentiment analysis results, the customer attention $CA_k(k = 1,2,\ldots,10)$ and customer satisfaction $CS_k(k = 1,2,\ldots,10)$ of CRs in the test set can be calculated using Equation (3) and Equation (4). Furthermore, the Kano classification results of CRs can be obtained using the method proposed in Figure 5 as shown in Table 5 and Figure 6.

Next, in order to assign reasonable weights to CRs, CR_6, CR_9, and CR_{10}, which are classified as indifferent qualities, are first directly eliminated, and then corresponding weight balance coefficients are set according to different Kano classifications. For must-be quality, $\alpha = \beta = 0.5$ is set; for one-dimensional quality, $\alpha = 0.7$, $\beta = 0.3$ is set; for attractive quality, $\alpha = 0.3$, $\beta = 0.7$ is set. Then the weight index WI_k and the final weight distribution results of CRs can be obtained, which are listed in Table 6.

Table 5 Kano classification of CRs.

	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6	CR_7	CR_8	CR_9	CR_{10}
CA_k	0.90	0.72	0.85	0.83	0.59	0.66	0.59	0.65	0.59	0.67
CS_k	0.50	0.41	0.50	0.57	0.45	0.42	0.46	0.49	0.34	0.32
Classification	M	O	M	M	A	I	A	A	I	I

Figure 6 Kano classification of CRs.

Table 6 Weight allocation of CRs.

	CR_1	CR_2	CR_3	CR_4	CR_5	CR_7	CR_8
WI_k	0.700	0.628	0.676	0.699	0.495	0.700	0.628
w_k	0.165	0.148	0.160	0.165	0.117	0.118	0.127

Stage C. Construction of QFD model with social network group consensus.

Stage C. 1. Collection of QFD model elements. Based on the actual operation of cruise ships, this study simulates the selection of five core service departments, as introduced below. These departments represent the critical components of cruise service quality and collectively cover the main dimensions of customer experience. The five service departments above form the QFD decision making group $E = \{e_1, e_2, \ldots, e_5\}$, dedicated to improving the quality of cruise ship services.

1) Food and Beverage Department (e_1): Responsible for dining services and meal experiences.

2) Cabin Department (e_2): Responsible for cabin cleaning, maintenance, and accommodation experience.

3) Entertainment Department (e_3): Responsible for organizing and implementing entertainment programs and activities.

4) Customer Service Department (e_4): Responsible for customer care, complaint handling, inquiry services, and other related matters.

5) Operations Management Department (e_5): Responsible for embarkation and disembarkation processes, as well as matters related to ship layout.

First, social network analysis is conducted among the five departments to identify their trust relationships. Drawing on SNA, it is assumed that reciprocal trust relationships exist among the departments. Each department express its trust toward others based on prior collaboration, resource dependencies, and willingness to share information. Specifically, based on the trust function introduced in section 3.3, the social trust matrix (STM) among $E = \{e_1, e_2, \ldots, e_5\}$ is represented as follows:

$$STM = \begin{pmatrix} & e_1 & e_2 & e_3 & e_4 & e_5 \\ e_1 & - & (0.6,0.2) & (0.5,0.1) & (0.7,0.3) & (0.8,0.1) \\ e_2 & (0.7,0.1) & - & (0.4,0.1) & (0.7,0.2) & (0.6,0.2) \\ e_3 & (0.4,0.3) & (0.6,0.2) & - & (0.5,0.2) & (0.8,0.1) \\ e_4 & (0.8,0.1) & (0.9,0.1) & (0.8,0.2) & - & (0.8,0.2) \\ e_5 & (0.7,0.1) & (0.7,0.2) & (0.6,0.1) & (0.5,0.3) & - \end{pmatrix} \tag{18}$$

Next, experts from each department determine the corresponding cruise service QCs for the 7 key CRs from their own professional perspectives, the results are shown in Table 7. And experts are also required to evaluate the relationship between CRs and cruise service QCs to construct the corresponding evaluation matrices, which are represented as follows:

$$F_1 = \begin{pmatrix} 1 & 0.3 & 0.2 & 0.7 & 0.2 \\ 0.8 & 0.8 & 0.3 & 0.5 & 0.3 \\ 0.8 & 0.5 & 0.3 & 1 & 0 \\ 0.8 & 0.5 & 0.3 & 1 & 0 \\ 0.4 & 0.3 & 0.7 & 0.6 & 0.3 \\ 0.2 & 0.2 & 0.1 & 0.4 & 0.8 \\ 0.5 & 0.6 & 0.5 & 0.5 & 0.7 \\ 0.3 & 0.7 & 0.3 & 0.5 & 0.3 \end{pmatrix}; F_2 = \begin{pmatrix} 0.7 & 0.5 & 0.3 & 0.5 & 0.4 \\ 0.4 & 1 & 0.2 & 0.6 & 0.6 \\ 0.5 & 0.8 & 0.4 & 0.9 & 0.5 \\ 0.2 & 0.5 & 0.7 & 0.6 & 0.4 \\ 0.3 & 0.6 & 0.2 & 0.5 & 0.9 \\ 0.4 & 0.9 & 0.5 & 0.4 & 0.8 \\ 0.2 & 1 & 0.3 & 0.4 & 0.5 \end{pmatrix}; F_3 = \begin{pmatrix} 0.6 & 0.3 & 0.5 & 0.6 & 0.2 \\ 0.3 & 0.8 & 0.7 & 0.5 & 0.4 \\ 0.5 & 0.5 & 0.8 & 1 & 0.4 \\ 0.3 & 0.4 & 1 & 0.8 & 0.4 \\ 0.2 & 0.3 & 0.4 & 0.5 & 0.8 \\ 0.3 & 0.5 & 0.9 & 0.7 & 0.7 \\ 0.2 & 0.6 & 0.5 & 0.6 & 0.3 \end{pmatrix};$$

$$F_4 = \begin{pmatrix} 0.8 & 0.4 & 0.3 & 1 & 0.4 \\ 0.4 & 0.7 & 0.4 & 1 & 0.5 \\ 0.7 & 0.5 & 0.5 & 1 & 0.6 \\ 0.3 & 0.3 & 0.6 & 0.8 & 0.4 \\ 0.2 & 0.3 & 0.2 & 0.7 & 1 \\ 0.5 & 0.6 & 0.5 & 0.6 & 0.8 \\ 0.3 & 0.7 & 0.2 & 0.4 & 0.4 \end{pmatrix}; F_5 = \begin{pmatrix} 0.7 & 0.3 & 0.2 & 0.5 & 0.7 \\ 0.3 & 0.6 & 0.3 & 0.5 & 0.8 \\ 0.5 & 0.5 & 0.4 & 0.9 & 0.8 \\ 0.3 & 0.3 & 0.8 & 0.6 & 0.7 \\ 0.3 & 0.5 & 0.4 & 0.6 & 1 \\ 0.5 & 0.7 & 0.6 & 0.7 & 0.9 \\ 0.4 & 0.8 & 0.4 & 0.6 & 0.8 \end{pmatrix};$$

Table 7 Cruise service QCs and descriptions.

QCs	Description
Dining Services ($\boldsymbol{QC_1}$)	Enhancing the dining experience on the cruise ship by improving the quality, diversity, and personalization of dining services.
Cabin and Cleanliness Management ($\boldsymbol{QC_2}$)	This dimension is dedicated to ensuring both the comfort and hygiene of the cabin environment.
Entertainment and Recreation ($\boldsymbol{QC_3}$	Enriching the variety and appeal of entertainment activities to meet the diverse needs of customers
Customer Care and Staff Interaction($\boldsymbol{QC_4}$)	Enhancement of service quality through comprehensive staff training to respond swiftly and professionally to customer needs.
Travel Process Efficiency($\boldsymbol{QC_5}$)	Optimizing the processes associated with embarkation, disembarkation, and other journey-related operations.

Stage C. 2. Construction of consensus mechanism. Due to the different responsibilities of each service department, there may be inconsistencies in the relationship evaluation matrices they provide. For example, the Food and Beverage Department is more concerned with dining services, while the Cabin Department focuses more on cabin and cleanliness management. In order to obtain reasonable and acceptable decision making results for all departments, it is necessary to promote consultation and communication between departments through a CRP.

From Equation (9), the initial bilateral consensus level matrix (BCLM) of five departments is obtained as follows:

$$BCLM = \begin{pmatrix} & e_1 & e_2 & e_3 & e_4 & e_5 \\ e_1 & - & 0.840 & 0.849 & 0.874 & 0.820 \\ e_2 & 0.840 & - & 0.814 & 0.857 & 0.854 \\ e_3 & 0.849 & 0.814 & - & 0.837 & 0.823 \\ e_4 & 0.874 & 0.857 & 0.837 & - & 0.843 \\ e_5 & 0.820 & 0.854 & 0.823 & 0.843 & - \end{pmatrix} \quad (19)$$

Furthermore, the general consensus level $GCL = 0.84$ is calculated. Without loss of generality, the consensus threshold $\gamma = 0.9$ is predefined, as a relatively high level of agreement is required in the context of multi-department decision making for cruise service management to ensure rationality and acceptability of the final results. Since $GCL < \gamma$, then the bilateral interaction consensus mechanism is implemented to promote consensus.

In each round of interaction, the two departments with the greatest conflict are first identified, and then the collaboration index between them is calculated based on their mutual consensus and trust levels by Equation (12). Next, the proposed bilateral interaction consensus model (13) is applied to realize the negotiation and opinion adjustments between experts. Let T denote the interaction round, after calculation, it is found that three rounds of interaction are required to achieve final group consensus, with Table 8 providing the indicators of each round of bilateral interaction.

Table 8 Indicators of each round of bilateral interaction.

Interaction round	Negotiating department	Collaboration index	Interaction parameters	GCL
$\boldsymbol{T=1}$	e_2, e_3	$CI_{2\to3}=0.728$; $CI_{3\to2}=0.755$	$\eta_2=0.218$; $\eta_3=0.226$	0.86
$\boldsymbol{T=2}$	e_1, e_5	$CI_{1\to5}=0.835$; $CI_{5\to1}=0.810$	$\eta_1=0.250$; $\eta_5=0.243$	0.88
$\boldsymbol{T=3}$	e_3, e_4	$CI_{3\to4}=0.748$; $CI_{4\to3}=0.829$	$\eta_3=0.224$; $\eta_4=0.249$	0.90

Stage C. 3. Prioritization of cruise service QCs. To assign appropriate weight to each department, the in-centrality degree of $E=\{e_1,e_2,\dots,e_5\}$ is calculated using Equation (14), and the weight vector of departments can be obtained by Equation (15): $\pi_1=0.202, \pi_2=0.205, \pi_3=0.195, \pi_4=0.182, \pi_5=0.215$.

Then, the consensus-based collective opinion can be obtained by aggregating the evaluation matrix of each department by Equation (16). Integrated with the weights of CRs $w_1=0.165, w_2=0.148, w_3=0.160, w_4=0.165, w_5=0.117, w_6=0.118, w_9=0.127$, the final score of cruise service QCs is derived by Equation (17): $FS_1=0.4516, FS_2=0.5441, FS_3=0.4518, FS_4=0.6575, FS_5=0.5574$, and the final prioritization of QCs is $QC_4>QC_5>QC_2>QC_3>QC_1$.

5.3 Suggestions for improving cruise service quality

Based on the application and calculation results of the proposed QFD model on the cruise ship Norwegian Breakaway, this section presents specific suggestions for improving cruise service quality, with the aim of better meeting customer requirements and improve overall service levels.

On the one hand, the following insights can be obtained from the Kano classification results of the obtained CRs:

(1) Must-be quality represents the fundamental requirements that customers regard as indispensable. In the context of cruise services, Dining Experience (CR_1), Staff and Customer Service(CR_3), and Onboard Activities and Entertainment (CR_4) all fall into this category. When these basic services fail to meet expectations, customer dissatisfaction is likely to occur. From a management perspective, it is essential to ensure that dining options are diverse and consistently high in quality, to provide regular training for staff to enhance their service professionalism and maintain stable service levels, and to offer a variety of entertainment activities to meet the needs of customers across different age groups and interests.

(2) Cleanliness and maintenance (CR_2) is evaluated by customers as one-dimensional quality, as customers have clear expectations regarding the cleanliness of the environment and the condition of facilities. Although the attention to this service is high, the existing service fails to fully meet customer expectations, resulting in relatively low satisfaction. From a management perspective, enterprises should establish strict cleanliness standards and maintenance plans, conducting regular environmental inspections and equipment upkeep to ensure all indicators meet the required standards, then the customer satisfaction can be effectively improved.

(3) Attractive quality refers to service attributes that exceed basic customer expectations and provide additional positive experiences. In the context of cruise services, the Embarkation and Disembarkation (CR_5), Ship Layout and Facilities (CR_7), and Cabin Experience (CR_8) fall into this category. These services typically do not attract high customer attention, but when optimized, they can significantly

enhance the overall customer experience. From a management perspective, companies should streamline embarkation and disembarkation processes, optimize ship layout, enhance facility configurations, and improve cabin environments and functionality. Such innovative improvements can create a differentiated advantage in the market and boosting customer loyalty.

(4) Indifferent quality refers to service attributes that receive low customer attention, where changes in service performance do not significantly impact overall satisfaction. In the context of cruise services, Excursions and Port Visits (CR_6), Communication and Organization (CR_9), and Pricing and Value for Money (CR_{10}) fall into this category. From a management perspective, cruise enterprises should ensure that these services meet basic standards and avoid significant shortcomings, but do not require excessive resource allocation.

On the other hand, from the final prioritization of cruise service QCs, the following suggestions can be obtained:

The ranking results indicate that Customer Care and Staff Interaction (QC_4) is the most critical factor influencing customer satisfaction. This may be because it is closely linked to multiple CRs, making it a central element of the overall service experience. To this end, cruise operators should enhance staff training, establish effective feedback mechanisms, and optimize personnel allocation, particularly during peak hours, to maintain consistent service quality. Following closely, Travel Process Efficiency (QC_5) ranks second, highlighting customers¡¯ expectations for a smooth boarding, disembarkation, and overall travel process. Lengthy procedures and extended waiting times can negatively affect satisfaction. Therefore, cruise operators should streamline processes by implementing digital check-in systems and intelligent queue management, and providing clear guidance to improve the travel experience.

Cabin and Cleanliness Management (QC_2) ranks third, emphasizing the importance of maintaining sanitation and facilities, and shortcomings in cleanliness can lead to dissatisfaction. Entertainment and Recreation (QC_3) and Dining Services (QC_1) rank lower in importance. This is because they are associated with more specific CRs, making their overall impact on satisfaction less significant compared to broader service elements. Enterprises should maintain a stable operation in these aspects and ensure the quality of basic services.

In summary, the proposed framework can be applied by different stakeholder groups in various aspects of decision making. Specifically, service designers may use the Kano-based weighting of CRs to identify and refine service attributes that most affect customer perceptions, thus supporting more targeted service design. Cruise operators can leverage the prioritization of service QCs to allocate resources more effectively and to design service processes that align with customer expectations. Meanwhile, customer experience managers can utilize the outputs of sentiment analysis and requirement extraction to monitor passenger satisfaction in real time and implement timely adjustments to enhance overall service outcomes. In this way, the framework not only contributes to theoretical understanding but also provides actionable guidance for stakeholders in the cruise industry.

6 Analysis and discussion

In this section, a systematic analysis and discussion of the proposed QFD model for cruise service improvement is provided. First, a sensitivity analysis of the relevant parameters in the proposed method is conducted to further reveal its characteristics and rationality; Then, a robustness analysis is performed to verify the stability of the proposed consensus mechanism. Finally, a quantitative comparison is included to validate the accuracy and effectiveness of the proposed prompt-based LLM approach, and a theoretical comparison is presented between the proposed QFD framework and existing research to demonstrate the advantages of this paper.

6.1 Sensitivity analysis

6.1.1 Sensitivity analysis of weight balance coefficients

In Section 4.2.2, weight balance coefficients α and β are set based on the different categories of Kano model to balance the impact of customer attention and customer satisfaction on the weight index, thereby assigning appropriate weights to the identified CRs. In this section, to verify the rationality and effectiveness of the proposed weight distribution method, balance coefficients α is varied between [0.1,0.9] with a step size of 0.1, and a sensitivity analysis is conducted to examine the effects of α on the CRs' weight index, weight index change rate, and normalized weight contribution. The results are shown in the Figures 7(a)- 7(c).

Figure 7(a) illustrates the effect of variations in α on the weight index of CRs. It can be observed that as α increases, the weight index WI_i becomes higher. This is because α reflects the degree of emphasis for customer attention. Since the value of customer attention is higher in numerical terms, it exerts a stronger influence on the weight of CRs. Figure 7(b) demonstrates how the relative change rate (RCR) of weight index varies with α, in which $RCR = \frac{WI_i - WI_{baseline}}{WI_{baseline}}$. When $\alpha = 0.5$, the weight change remains relatively stable; however, as α deviates from 0.5, the weight change rate increases significantly, indicating that extreme biases toward customer attention or satisfaction result in greater sensitivity in weight changes, which may impact decision making stability. Figure 7(c) reveals that when α is in the range of [0,0.5], the changes in the normalized weights are relatively small, whereas when α is within (0.5,1], the weight contributions show more noticeable changes.

From the observations above, α determines the degree of emphasis on customer attention, and moderate values of α help maintain the stability of weight allocation. This finding provides practical guidance for selecting weight balance coefficients in real applications. When customer attention is the primary driver, such as in one-dimensional quality, a relatively higher α may be adopted to reflect stronger attention effects. When customer satisfaction is dominant, such as in attractive quality, a smaller α is preferred to enhance satisfaction contribution. For must-be quality, a balanced α value around the central region is suitable. In addition, extreme values (e.g., α =0.1 or 0.9) should be avoided, as they lead to large weight fluctuations and may undermine stability in decision making. Therefore, the sensitivity analysis not only reveals the impact of α but also provides empirical evidence supporting reasonable parameter selection.

(a) Weight index change with different α.

(b) Weight index change rate with different α.

(c) Normalized weight comparison.

Figure 7 Sensitivity analysis of weight balance coefficients.

6.1.2 Sensitivity analysis of group consensus mechanism

In the constructed bilateral interaction consensus model (13), the decision variables η_p and η_q are primarily influenced by the parameter θ and the collaboration index (CI). CI is affected by the level of consensus and trust between individuals, while θ reflects the impact of CI on the interaction parameters. Figure 8(a) illustrates how the collaboration index is influenced by the bilateral consensus level and trust score, while Figure 8(b) depicts the evolution process of GCL as θ varies, with $\theta = \{0.1,0.2,0.3,0.4,0.5\}$ is set.

From Figure 8(a), it can be observed that CI exhibits a nonlinear upward trend as bilateral consensus level and trust score increase, which aligns with the original intention of constructing CI, and higher trust and consensus levels jointly promote the rapid improvement of the CI, and both are indispensable. From Figure 8(b), it is evident that as θ increases, the number of interaction rounds required to reach group consensus decreases, indicating an improvement in consensus efficiency. This is because within the proposed consensus model, a higher θ implies that experts are willing to make greater concessions under the same CI, thereby accelerating the process of reaching consensus.

(a) CI change with consensus and trust levels.

(b) GCL evolution process with different θ.

Figure 8 Sensitivity analysis of CI and consensus reaching process.

6.2 Robustness analysis

To further assess the stability of the proposed bilateral interaction consensus mechanism, this section conducts a robustness analysis by introducing controlled perturbations to the initial expert evaluation matrices. The aim is to verify whether the consensus outcome remains more stable than individual expert assessments when the preference information is subject to noise or uncertainty.

Starting from the five original matrices, random noise sampled from a uniform distribution is added to each evaluation value. The perturbation magnitude is set to vary incrementally from 0.1 to 1.0, with levels defined as $\epsilon \in \{0.1, 0.2, \ldots, 1.0\}$. The perturbed matrices are generated using

$$\tilde{F} = clip(F + U[-\epsilon, \epsilon], 0, 1),$$

where the clipping function ensures values remain within the original evaluation scale. Each perturbation level is repeated ten times to obtain statistically stable estimates. For each experiment, both individual expert rankings and the consensus ranking are computed before and after perturbation. Individual ranking is obtained by averaging each alternative's row values, and the consensus ranking is obtained by running the bilateral interaction mechanism on both F and $\tilde{F}$.

The impact of perturbation on ranking results is quantified using the number of Kendall discordant pairs, defined as cases in which the relative ordering of a pair of alternatives is reversed after perturbation. A larger number of discordant pairs indicates lower stability and stronger sensitivity to data fluctuation. This measure is computed for both individual experts and the final consensus ranking, and the results are shown in Figure 9, from which the following observations can be concluded:

(1) As the perturbation intensity increases, both the average individual ranking changes and the consensus ranking changes show an upward trend, indicating that noise indeed affects evaluation stability under uncertain information.

(2) Across all perturbation levels, the consensus ranking change consistently remains lower than the average individual ranking change, demonstrating that the consensus mechanism provides greater stability and robustness. This suggests that the proposed bilateral interaction model can effectively improve the resilience and reliability of group decision making outcomes.

Figure 9 Average number of Kendall discordant pairs under different noise levels

6.3 Comparative analysis

6.3.1 Quantitative comparison

To validate the accuracy and effectiveness of the proposed prompt-based LLM approach in customer requirement extraction, this study further compares it with three widely used traditional methods: TF-IDF, LDA, and BERT.

Term Frequency–Inverse Document Frequency (TF-IDF) is a classical method for weighting textual features. Its basic principle is to balance the frequency of a word within a single document (term frequency, TF) against its prevalence across the entire corpus (inverse document frequency, IDF), thereby highlighting words that are distinctive in individual reviews but not overly common in the dataset as a whole. In this study, TF-IDF is applied to 500 cruise reviews to extract the top 100 candidate terms, with their corresponding document frequency, IDF values, and mean weights calculated. The results for the top 5 terms are presented in Table 9.

Latent Dirichlet Allocation (LDA) is a probabilistic topic modeling method that assumes each document is composed of a mixture of latent topics, and each topic is represented by a distribution over words. Through this generative process, LDA infers hidden structures in text and assigns words to topics based on their co-occurrence patterns. In this study, LDA is applied to a corpus of 500 cruise reviews, with the model set to generate 10 latent topics. The results of the first five topics are presented, with each topic shown its top-ranked keywords and their associated weights, the results are summarized in Table 10.

Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained language model that captures contextualized semantic representations of words and phrases. Unlike purely statistical approaches such as TF-IDF and LDA, BERT leverages deep neural network architectures and large-scale pretraining to identify semantically meaningful expressions. In this study, a pre-trained BERT model is employed in an unsupervised setting through the KeyBERT framework to extract candidate attributes directly from the corpus of 500 cruise reviews. The model produced a ranked list of key terms, each associated with a relevance score that reflects its semantic importance within the corpus. In addition, the number of reviews in which each term appeared and the corresponding coverage rate across the dataset are recorded. The results for the top-ranked terms are shown in Table 11.

Table 9 Top 5 candidate terms extracted by TF-IDF.

Term	Document Frequency	Inverse Document Frequency (IDF)	Mean TF-IDF weight
dining	227	1.7892	0.0416
restaurants	190	1.9663	0.0415
port	188	1.9768	0.0449
buffet	188	1.9768	0.0416
pool	186	1.9874	0.0448

Table 10 Top keywords and weights for the first five topics by LDA.

Topic ID	Keyword and their weight
1	room(0.0104), port(0.0070), area(0.0046), bar(0.0038), luggage(0.0038)…
2	port(0.0070), staff(0.0056), room(0.0056), excursion(0.0049), entertainment(0.0036)…
3	room(0.0104), service(0.0068), buffet(0.0061), staff(0.0045), dinner(0.0045)…
4	service(0.0068), cabin(0.0058), entertainment(0.0057), buffet(0.0061),pool(0.0052)…
5	staff(0.0045), fun(0.0068), night(0.0060), pool(0.0052), drinks(0.0042)…

Table 11 Top 5 candidate terms extracted by BERT.

Term	Relevance score	Hit reviews
boarding process	0.3702	10

the customer	0.1844	10
same menu	0.1358	11
vibe	0.1272	15
rush	0.1217	9

From the results above, it is evident that TF-IDF, LDA, and BERT each produced lists of candidate terms or phrases, rather than fully coherent themes. While these outputs provide useful indicators of customer concerns, they cannot directly represent higher-level service attributes. Consequently, a manual semantic clustering step is required to consolidate the candidate terms into ten broader customer requirement themes. The summarized results are presented in Table 12, which compares the ten themes derived from the three methods.

From a theoretical perspective, the three baseline methods each have inherent limitations. TF-IDF relies on word frequency statistics and therefore cannot capture contextual meaning, often highlighting common or fragmented terms that lack semantic coherence. LDA, as a probabilistic topic model, is able to group words into latent topics, but the topics frequently overlap and require substantial human interpretation to become meaningful service attributes. BERT-based keyword extraction leverages semantic embeddings and produces more contextually relevant terms, but it still remains limited to surface-level correlations and cannot generate structured, domain-specific themes without external guidance. By contrast, the LLM with prompt engineering can comprehend review texts in context, inferring higher-level attributes, and directly aligning the extracted results with interpretable service dimensions, thereby reducing the gap between raw textual data and actionable managerial insights.

Table 12 Comparison of manually summarized themes from TF-IDF, LDA, and BERT.

Methods	Derived themes
TF-IDF	Cabin & Onboard Accommodation, Dining & Food Options, Staff & Crew Interaction, Entertainment & Social Activities, Ports & Shore Excursions, Booking & Reservation Experience, Customer Complaints & Negative Issues, Facilities & Onboard Environment, Cruise Line & Brand Mentions, Timing & Scheduling Concerns
LDA	Accommodation & Embarkation experience, Dining & entertainment facilities, Nightlife &social activities, Dining & bar experience, Overall service & dining experience, Entertainment shows & overall experience, Accommodation & evening entertainment, Beverage & overall service
BERT	Dining Experience, Cabin & Accommodation, Embarkation & Disembarkation Process, Excursions & Destinations, Onboard Entertainment &Activities, Staff & Customer Service, Pricing & Value for Money, Ship Facilities & Space, Communication & Organization, Overall Cruise Experience & Impressions

To further evaluate the accuracy of attribute extraction, a quantitative validation by comparing the outputs of different methods with the manually annotated set of 100 cruise reviews is conducted. Two widely used evaluation metrics are employed: the F1-score and the Jaccard similarity coefficient.

The F1-score is the harmonic mean of precision and recall, which balances the trade-off between the proportion of correctly predicted attributes among all predictions and the proportion of correctly predicted attributes among all true labels. Formally,

$$Precision = \frac{TP}{TP+FP}, \;\; Recall = \frac{TP}{TP+FN};$$

$$F1 - \text{score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}.$$

The Jaccard similarity coefficient measures the overlap between the predicted attribute set A and the gold standard attribute set B. A higher Jaccard score indicates a greater consistency between the predicted and annotated attributes. It is defined as:

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Using these two metrics, we evaluate three baseline methods (TF-IDF, LDA, and BERT) and the proposed LLM-based prompt engineering approach against the human-annotated dataset. The results are summarized in Table 13. The results show that TF-IDF and LDA tend to extract high-frequency or statistically significant words but lack semantic understanding, leading to lower performance in F1-score and Jaccard similarity. BERT captures semantic associations better leveraging word embeddings but remains limited without contextual reasoning. In contrast, the proposed LLM-based prompt engineering approach significantly outperforms the three traditional methods on both evaluation metrics. This method is able to comprehensively understand the context and extract attributes that better reflect customers' real needs, while its semantic generalization shows a high level of consistency with the human annotations. These results demonstrate that LLMs exhibit stronger generalization capacity and semantic understanding when handling unstructured review data.

Table 13 Performance comparison of different methods against human annotation

Methods	F1-score	Jaccard
TF-IDF	0.4204	0.2700
LDA	0.5340	0.3668
BERT	0.5583	0.3590
LLM	0.6750	0.4667

Furthermore, the four approaches are evaluated to compare the computational overhead on the same cruise review dataset. A performance monitoring module into each script is embedded, using time.perf_counter() to measure total runtime, tracemalloc to capture peak Python memory allocation, and psutil to sample the resident set size (RSS) as the peak physical memory usage. In this experiment, the three traditional methods are applied to the full set of 500 reviews to ensure stable results based on statistical features and topic modeling. In contrast, the LLM relies on semantic understanding rather than large sample sizes and is subject to API token limitations; therefore, it is conducted in batches and reported as the equivalent of processing 100 reviews for comparison. The results of the four methods are summarized in Table 14.

The comparison shows that the traditional methods differ substantially in efficiency and memory consumption. TF-IDF has the shortest runtime, LDA requires more time but remains within an acceptable range. In contrast, BERT demands much more runtime and memory than the first two methods, mainly because it needs to load and run a deep language model to generate embeddings, which significantly increases the local computational and storage burden. LLM approach takes longer overall and is slower than all three traditional methods, but its local resource usage is extremely low. This is because the main computation is handled by remote servers, and the local machine only needs to construct prompts and communicate with the API. As a result, the local memory footprint remains minimal, though runtime is influenced by network latency. Overall, while the LLM method is not the most efficient in terms of speed, it offers clear advantages in reducing local resource consumption and in providing stronger semantic abstraction, making it suitable for scenarios where computational resources are limited or high-quality semantic summarization is required.

Table 14 Comparison of runtime and memory usage across different methods

Methods	Runtime (s)	Peak python memory (MB)	Peak physical memory (MB)
TF-IDF	2.496	40.41	190.11
LDA	14.605	20.18	190.11
BERT	77.139	3141.29	3678.05
LLM	156.62	9.20	108.60

6.3.2 Theoretical comparison

(1) Comparison with existing research on QFD

In order to further illustrate the superiority of the proposed QFD framework, a theoretical comparison is presented between the proposed method and some existing research on QFD from five aspects, the results are shown in Table 15.

1) In terms of CRs extraction, existing studies primarily adopt two approaches: (1) obtaining CRs directly through surveys and interviews, or (2) utilizing traditional text processing techniques to extract information from online reviews and other data sources. In contrast, this study leverages the advanced text analysis and processing capabilities of LLMs to extract key customer needs and conduct sentiment analysis from online review data. This approach enables an in-depth analysis and provides more authentic and reliability analytical results of CRs extraction.

2) In terms of CRs weight allocation, most existing studies employ well-established weighting techniques or consider enterprise life cycle to determine CRs' weights. However, these approaches largely rely on expert-provided evaluation data, which may introduce subjective bias or cognitive limitations. In contrast, this study derives CRs' weights based on the analysis of online review data. Specifically, the weight balance coefficients are assigned to different Kano classifications, ensuring a more objective and rational weight determination. By reducing reliance on subjective expert judgments, this approach enhances the objectivity and reliability of weight allocation.

3) In terms of the QFD decision making process, some studies have developed consensus-based QFD decision approaches to effectively integrate the collective wisdom of experts, thereby achieving more acceptable and reliable decision outcomes. However, these methods typically rely on traditional centralized interaction mechanisms, which may be insufficient in identifying and resolving critical conflicts among experts. To address this limitation, this study proposes a bilateral interaction consensus strategy designed to dynamically detect and manage disagreements among expert opinions. This approach enhances the efficiency of QFD group decision making and facilitates a more effective CRP.

4) In terms of constructing social trust network of experts, although most studies have not fully considered the potential impact of trust relationships among experts on the QFD decision making process, a few studies have explored this aspect. For example, Liu at al. (2022) utilized SNA to generate the weights of departments, but ignores the role of trust relationships in consensus building. In this study, SNA is employed to investigate

the social trust matrix among experts, aiming to uncover the importance weight of each expert within the group and promote more effective interactions among experts, thereby fostering the achievement of consensus.

5) Finally, in terms of application domains, most existing QFD research focus primarily on the design and optimization of industrial or electronic products, with relatively limited research on the design and improvement of service-oriented products. The cruise industry, as one of the most dynamic sectors of the global economy, requires ongoing exploration of service demands and continuous improvement of service quality. This study focuses on this topic to provide actionable solutions and guidance for enhancing service quality in the cruise industry.

Table 15 Comparison between the proposed QFD model and existing research.

References	CRs extraction	CRs weighting	Is QFD decision consensus-based?	Is social trust between experts considered?	Application domain
Li et al. **(2019)**	Collected from open design	Based on importance	No	No	Smartphone design
Fang et al. **(2020)**	Questionnaires, interviews, and focus groups	Integrate subjective and objective weights	No	No	Rotor service design
Liu et al. **(2019)**	Market investigation and interviews with experts	Based on subjective importance	No	No	Electric vehicle design
Wu et al. **(2021)**	Survey by the marketing department	Considering enterprise life cycle	No	No	Chinese-commerce service design
Chen et al. **(2021)**	Given by experts	DEMATEL and extended MULTIMOORA	No	No	CNC machine tool design
Liu et al. **(2022)**	Extracted from online reviews based on word frequency	Considering enterprise life cycle	Yes	Yes	Chinese new e-commerce mode
Gai et al. **(2024a)**	Extracted from online reviews based on word frequency	CRP of decision expert evaluation	Yes	Yes	Product design of e-commerce platforms
Wang et al. **(2024a)**	Interview with experts	Interview with experts	Yes	No	Energy storage technology development
Yang et al. **(2024)**	Identified by QFD team	By House of Quality-based relationship matrix	Yes	No	Car-sharing product- service system
This study	Extracted from online reviews using LLMs	Kano model-based weight balance coefficients	Yes	Yes	Cruise service quality improvement

(2) Comparison with existing research on service quality and customer satisfaction

To position the proposed methodology within the broader literature, a comparative review is conducted against the representative studies in the hospitality and tourism domain. The selected works cover a variety of approaches, ranging from survey-based SERVQUAL applications and structural equation modeling (SEM) to bibliometric mapping and fuzzy-set qualitative comparative analysis (fsQCA). For consistency, each study is compared along three dimensions: research context, methodology, and strengths and limitations. The results of this comparison are summarized in Table 16, some observations can be obtained as follows:

1) Most existing studies in hospitality and tourism rely on survey-based approaches, often grounded in SERVQUAL or similar frameworks. These methods provide strong theoretical foundations and allow for structured, quantifiable analysis, but they are typically limited by small samples, narrow geographic focus, and an inability to capture dynamic or unstructured customer needs.

2) In recent years, newer approaches have emerged. Bibliometric analyses map research trends and themes, fsQCA reveals causal complexity, and studies contrasting AI and human service quality highlight digital transformation in service contexts. Each offers valuable perspectives, yet they also face constraints: bibliometrics lack primary data, fsQCA has limited generalizability, and AI–human comparisons still depend on surveys and are prone to bias.

3) The proposed LLM-based methodology complements these approaches by shifting from structured questionnaires to unstructured online reviews. By integrating semantic extraction, Kano classification, and a consensus mechanism within a QFD framework, it identifies both explicit and implicit customer needs while reducing expert subjectivity. Although dependent on API access and requiring further validation across contexts, this method offers a data-driven, scalable alternative that better reflects the dynamic nature of customer feedback in hospitality.

Table 16 Comparison between the proposed QFD model and existing research.

References	Research context	Methodology	Strengths and limitations
Ali et al. **(2021)**	Hotel service quality and customer satisfaction in Erbil	Survey of 111 customers, SERVQUAL dimensions, factor and regression analysis	Strengths: strong theoretical grounding, structured measurement, practical implications; Limitations: small sample, single region, limited ability to capture emerging needs
PJ et al. **(2023)**	Service quality and customer satisfaction in hospitality, leisure, sport, and tourism	Bibliometric analysis of 740 WoS articles using Bibliometrix	Strengths: comprehensive mapping of research trends and themes; Limitations: relies only on WoS, lacks primary data and empirical validation
Chandra et al. **(2021)**	Hotel service quality and brand image in relation to satisfaction and loyalty	Survey of 120 customers, Structural Equation Modeling (SEM)	Strengths: integrates service quality and brand image, tests causal paths; Limitations: small sample, single hotel, reliance on self-reported data
Perdomo-Verdecia et al. **(2024)**	Hotel service quality configurations and customer satisfaction	fsQCA applied to 30k+ survey responses, 42 cases analyzed	Strengths: captures causal complexity and multiple pathways; Limitations: single hotel, survey-based bias, limited generalizability
Sardesai et al. **(2024)**	AI service quality vs. human service quality in Indian hotels	Survey of 138 customers, SEM with mediation analysis	Strengths: timely focus on AI vs. human service, rigorous SEM analysis; Limitations: limited sample and region, subjective bias from surveys

Anas et al. **(2024)**	Hotel service quality and customer satisfaction in Makassar City	Survey of 680 customers, SERVQUAL dimensions, multiple regression analysis	Strengths: large sample, robust statistical tests, practical managerial insights; Limitations: single hotel, cross-sectional design, restricted to SERVQUAL dimensions
This study	Cruise service quality and customer needs extraction	LLMs for need extraction and sentiment analysis; Kano model; consensus mechanism integrated into QFD	Strengths: innovative integration of LLMs with Kano and consensus methods, captures implicit needs, reduces expert bias, scalable to broader services; Limitations: Kano classification partly subjective, expert trust network relatively static

7 Conclusion

This study proposes a systematic QFD framework that integrates LLMs, the Kano model, and a social network-based group consensus mechanism, providing practical managerial insights for improving service quality in the cruise industry. First, by leveraging LLM GPT-4o combined with prompt engineering, this study extracts key CRs and conducts sentiment analysis on cruise review data collected from CruiseCritic, establishing a solid data foundation for subsequent model construction. Next, based on sentiment analysis results, the study classifies CRs using the Kano model and introduces weight balance coefficients to assign appropriate weights to CRs, ensuring a more objective and accurate assessment of their importance. Finally, a social network-based bilateral interaction consensus mechanism is developed to identify and resolve critical disagreements between QFD decision experts, which lead to more reliable and actionable insights for enhancing cruise service quality. Additionally, a case study of Norwegian Breakaway is conducted, some sensitivity analysis and comparative analysis are also provided to further highlight the characteristics and advantages of the proposed approach.

The theoretical and practical implications of this article can be summarized as follows:

- This study advances the theoretical foundation of QFD by integrating LLMs, the Kano model, and a social network-based consensus mechanism, offering a novel methodological perspective for QFD research. Traditionally, CRs in QFD are extracted through expert interviews and surveys, which are often subjective and resource-intensive. By leveraging LLMs for automated CR extraction and sentiment analysis, this study introduces a data-driven approach that enhances the objectivity and efficiency of CR identification. Furthermore, by incorporating sentiment analysis data to guide Kano classification and introducing weight balance coefficients, this study enhances the objectivity of CR importance evaluation, making the QFD framework more data-informed and scientifically grounded.

- This study provides practical guidance for cruise service management by establishing a data-driven QFD framework that enhances decision making accuracy and efficiency. By utilizing LLMs to automatically extract and analyze CRs from large-scale online reviews, cruise companies can gain deeper insights into passenger needs while reducing reliance on manual analysis. The refined Kano-based classification and weight allocation method ensure a more objective assessment of CRs, allowing businesses to prioritize service improvements based on real customer sentiment. Furthermore, the proposed consensus mechanism facilitates more effective expert decision making by resolving conflicts and improving agreement levels, ultimately leading to more stable and widely accepted QFD decisions. Beyond the cruise industry, this framework can be extended to other service sectors, such as hospitality, aviation, and tourism, providing a scalable approach for optimizing service quality based on customer-driven insights.

There are certain limitations that can addressed in future research:

(1) First, although the Kano model combined with weight balance coefficients provides an effective methods for the weight allocation of customer requirements, the process of defining category boundaries is still influenced by subjective judgment. This subjectivity may affect the consistency and reliability of classification outcomes. Future studies could reduce this limitation by introducing adaptive boundary-setting rules or

integrating machine learning–based clustering methods that automatically detect shifts in customer preferences from large-scale review data.

(2) Second, the consensus mechanism designed in this study is based on a predefined social network of expert trust relationships. While this approach helps to model interpersonal influence, it does not fully capture the dynamic and evolving nature of expert interactions over time. In practice, expert trust and influence may change with ongoing discussions or exposure to new evidence. Future research could therefore enhance the model by incorporating real-time feedback loops or dynamic network updating strategies, enabling the consensus mechanism to better reflect changing interaction patterns and improve its adaptability and accuracy.

(3) Furthermore, it should be acknowledged that this study is based on a single-case analysis, which limits the generalizability of the findings. Nevertheless, the proposed framework demonstrates scalability and can be applied to other service contexts that rely on large-scale online reviews, such as different cruise routes, hotels, tourist attractions, and e-commerce platforms. Future research may adopt multi-case and cross-industry validation to enhance the framework's generalizability and practical value.

(4) Finally, this study focuses on integrating LLM-based attribute extraction and sentiment analysis into an improved QFD framework, without systematically comparing it with current Transformer-based or LLM-based natural language processing (NLP) methods. Subsequent work will incorporate comparative analyses with state-of-the-art NLP benchmark models in specific task scenarios to more comprehensively evaluate the framework's performance.

Ahn, H. and Park, E., *Motivations for user satisfaction of mobile fitness applications: An analysis of user experience based on online review comments.* Humanities and Social Sciences Communications, 2023. **10**(1), 1-7.

Ali, B.J., Gardi, B., Othman, B.J., Ahmed, S.A., Ismael, N.B., Hamza, P.A., Aziz, H.M., Sabir, B.Y., Sorguli, S., and Anwar, G., *Hotel service quality: The impact of service quality on customer satisfaction in hospitality.* International Journal of Engineering, Business and Management, 2021. **5**(3), 14-28.

Anas, M., *Customer satisfaction in the hotel industry: A case study from Makassar City.* Journal of Economic Education and Entrepreneurship Studies, 2024. **5**(1), 135-150.

Bondy, J.A. and Murty, U.S.R., *Graph theory with applications*. Vol. 290. 1976: Macmillan London.

Cao, M., Chiclana, F., Liu, Y., Wu, J., and Herrera-Viedma, E., *A bilateral negotiation mechanism by dynamic harmony threshold for group consensus decision making.* Engineering Applications of Artificial Intelligence, 2024. **133**, 108225.

Cao, M., Gai, T., Wu, J., Chiclana, F., Zhang, Z., Dong, Y., Herrera-Viedma, E., and Herrera, F., *Social network group decision making: Characterization, taxonomy, challenges and future directions from an AI and LLMs perspective.* Information Fusion, 2025. **120**, 103107.

Castillo-Manzano, J.I., Castro-Nuño, M., and Pozo-Barajas, R., *Addicted to cruises? Key drivers of cruise ship loyalty behavior through an e-WOM approach.* International Journal of Contemporary Hospitality Management, 2022. **34**(1), 361-381.

Chandra, T. and Putra, R., *Service quality and brand image on customer satisfaction and customer loyalty at Pesonna Hotel Pekanbaru.* Journal of Applied Business and Technology, 2021. **2**(2), 142-153.

Chen, Y., Ran, Y., Huang, G., Xiao, L., and Zhang, G., *A new integrated MCDM approach for improving QFD based on DEMATEL and extended MULTIMOORA under uncertainty environment.* Applied Soft Computing, 2021. **105**, 107222.

Chrysafis, K.A., Papadopoulou, G.C., and Theotokas, I.N., *Measuring financial performance through operating business efficiency in the global cruise industry: A fuzzy benchmarking study on the "big three".* Tourism Management, 2024. **100**, 104830.

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A.S., Ceder, G., Persson, K.A., and Jain, A., *Structured information extraction from scientific text with large language models.* Nature communications, 2024. **15**(1), 1418.

Dong, Y., Ding, Z., and Kou, G., *Social network DeGroot model.* Springer Books, 2024.

Fang, H., Li, J., and Song, W., *A new method for quality function deployment based on rough cloud model theory.* IEEE transactions on engineering management, 2020. **69**(6), 2842-2856.

Gai, T., Wu, J., Liang, C., Cao, M., and Zhang, Z., *A quality function deployment model by social network and group decision making: Application to product design of e-commerce platforms.* Engineering Applications of Artificial Intelligence, 2024. **133**, 108509.

Gai, T., Wu, J., Chiclana, F., Cao, M., and Yager, R.R., *Dynamic compromise behavior driven bidirectional feedback mechanism for group consensus with overlapping communities in social network.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(10), 6149-6161.

Guo, W., Wang, H., Zhang, W.-G., Gong, Z., Xu, Y., and Słowiński, R., *Multi-dimensional multi-round minimum cost consensus models with iterative mechanisms involving reward and punishment measures.* Knowledge-Based Systems, 2024. **293**, 111710.

He, L., Wu, Z., Xiang, W., Goh, M., Xu, Z., Song, W., Ming, X., and Wu, X., *A novel Kano-QFD-DEMATEL approach to optimise the risk resilience solution for sustainable supply chain.* International journal of production research, 2021. **59**(6), 1714-1735.

Ji, F., Cao, Q., Li, H., Fujita, H., Liang, C., and Wu, J., *An online reviews-driven large-scale group decision making approach for evaluating user satisfaction of sharing accommodation.* Expert Systems with Applications, 2023. **213**, 118875.

Ji, F., Wu, J., Chiclana, F., Sun, Q., Liang, C., and Herrera-Viedma, E., *Supporting group cruise decisions with online collective wisdom: An integrated approach combining review helpfulness analysis and consensus in social networks.* Information Processing & Management, 2025. **62**(1), 103936.

Jiao, Y., Lau, Y.-y., and Gao, J., *Exploring the factors affecting cruise passengers' perceptions of value for money expressed in online reviews.* Humanities and Social Sciences Communications, 2024. **11**(1), 1-11.

Kano, N., Seraku, N., Takahashi, F., and Tsuji, S., *Attractive quality and must-be quality.* 1984.

Li, S., Tang, D., and Wang, Q., *Rating engineering characteristics in open design using a probabilistic language method based on fuzzy QFD.* Computers & Industrial Engineering, 2019. **135**, 348-358.

Liu, H.-C., Wu, S.-M., Wang, Z.-L., and Li, X.-Y., *A new method for quality function deployment with extended prospect theory under hesitant linguistic environment.* IEEE Transactions on Engineering Management, 2019. **68**(2), 442-451.

Liu, P., Zhang, K., Dong, X., and Wang, P., *A big data-Kano and SNA-CRP based QFD model: Application to product design under Chinese new E-commerce model.* IEEE Transactions on Engineering Management, 2022. **71**, 4246-4260.

Liu, P., Li, Y., and Wang, P., *Social trust-driven consensus reaching model for multiattribute group decision making: Exploring social trust network completeness.* IEEE Transactions on Fuzzy Systems, 2023. **31**(9), 3040-3054.

Ma, W., Ji, F., Liang, C., Sun, Q., and Wu, J., *A deep learning and large group consensus based cruise satisfaction evaluation model with online reviews.* Information Sciences, 2024. **676**, 120801.

Qorib, M., Oladunni, T., Denis, M., Ososanya, E., and Cotae, P., *Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset.* Expert Systems with Applications, 2023. **212**, 118715.

Perdomo-Verdecia, V., Garrido-Vega, P., and Sacristán-Díaz, M., *An fsQCA analysis of service quality for hotel customer satisfaction.* International Journal of Hospitality Management, 2024. **122**, 103793.

PJ, S., Singh, K., Kokkranikal, J., Bharadwaj, R., Rai, S., and Antony, J., *Service quality and customer satisfaction in hospitality, leisure, sport and tourism: An assessment of research in web of science.* Journal of Quality Assurance in Hospitality & Tourism, 2023. **24**(1), 24-50.

Sardesai, S., D'Souza, E., and Govekar, S., *Analysing the impacts of artificial intelligence service quality and human service quality on customer satisfaction and customer loyalty in the hospitality sector.* Turizam, 2024. **28**(1), 37-48.

Sun, X., Kwortnik, R., Xu, M., Lau, Y.-y., and Ni, R., *Shore excursions of cruise destinations: Product categories, resource allocation, and regional differentiation.* Journal of Destination Marketing & Management, 2021. **22**, 100660.

Sun, X., Cao, X., and Lau, Y.-y., *Exploring cruise tourists' sentiment expression pattern from online reviews: an insight into market positioning.* Tourism Management Perspectives, 2023. 49, 101195.

Suresh Kumar, K., Radha Mani, A., Ananth Kumar, T., Jalili, A., Gheisari, M., Malik, Y., Chen, H.-C., and Jahangir Moshayedi, A., *Sentiment analysis of short texts using SVMs and VSMs-based multiclass semantic classification.* Applied Artificial Intelligence, 2024. **38**(1), 2321555.

Tiutiu, M., Nemțeanu, S., Dabija, D.-C., and Pelau, C., *The impact of online customer service and store features on consumer experience and willingness to revisit their preferred online store.* Humanities and Social Sciences Communications, 2025. **12**(1), 1-13.

Venugopalan, M. and Gupta, D., *An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis.* Knowledge-based systems, 2022. **246**, 108668.

Victor, P., Cornelis, C., De Cock, M., and Da Silva, P.P., *Gradual trust and distrust in recommender systems.* Fuzzy Sets and Systems, 2009. **160**(10), 1367-1382.

Wan, B., Wu, P., Yeo, C.K., and Li, G., *Emotion-cognitive reasoning integrated BERT for sentiment analysis of online public opinions on emergencies.* Information Processing & Management, 2024. **61**(2), 103609.

Wang, J., Liu, H.-C., Zhang, J., Shi, H., and Zhang, Q.-Z., *New approach for quality function development based on cooperative game-based consensus mechanism and three-way decision method.* International Journal of Production Economics, 2024. **276**, 109380.

Wang, Q., Zhang, W., Li, J., Mai, F., and Ma, Z., *The devil is in the details! Effect of differentiated platform governance on online review manipulation.* Humanities and Social Sciences Communications, 2024. **11**(1), 1-14.

Wasserman, S., *Social network analysis: Methods and applications.* The Press Syndicate of the University of Cambridge, 1994.

Wu, H., Zhang, Z., Shi, S., Wu, Q., and Song, H., *Phrase dependency relational graph attention network for aspect-based sentiment analysis.* Knowledge-Based Systems, 2022. **236**, 107736.

Wu, T., Liu, X., Qin, J., and Herrera, F., *An interval type-2 fuzzy Kano-prospect-TOPSIS based QFD model: Application to Chinese e-commerce service design.* Applied Soft Computing, 2021. **111**, 107665.

Xu, Y., Ju, Y., Gong, Z., Sun, J., Dong, P., Ju, T., and Herrera-Viedma, E., *Improving consensus in social network group decision-making: Emphasizing overlapping subgroups and interactive behaviors.* Information Sciences, 2024. **679**, 121065.

Yang, Q., Chen, Z.-S., Zhu, J.-H., Martínez, L., Pedrycz, W., and Skibniewski, M.J., *Concept design evaluation of sustainable product–service systems: A QFD–TOPSIS integrated framework with basic uncertain linguistic information.* Group Decision and Negotiation, 2024. **33**(3), 469-511.

Yu, G., Zhang, Z., and Wu, J., *A stability analysis for the online retailing cyber security situation piecewise variable weight rating*

method. Applied Intelligence, 2025, **55**(18), 1144.
Zha, Q., Liang, H., Kou, G., Dong, Y., and Yu, S., *A feedback mechanism with bounded confidence-based optimization approach for consensus reaching in multiple attribute large-scale group decision-making.* IEEE Transactions on Computational Social Systems, 2019. **6**(5), 994-1006.
Zhang, Y., Yang, X., Xu, X., Gao, Z., Huang, Y., Mu, S., Feng, S., Wang, D., Zhang, Y., and Song, K., *Affective computing in the era of large language models: A survey from the nlp perspective.* arXiv preprint arXiv:2408.04638, 2024.
Zhou, X., Li, S., and Wei, C., *Consensus reaching process for group decision-making based on trust network and ordinal consensus measure.* Information Fusion, 2024. **101**, 101969.

Data Availability Statement

The data supporting the findings of this study are available within the supplementary files. Specifically, the following materials have been shared:

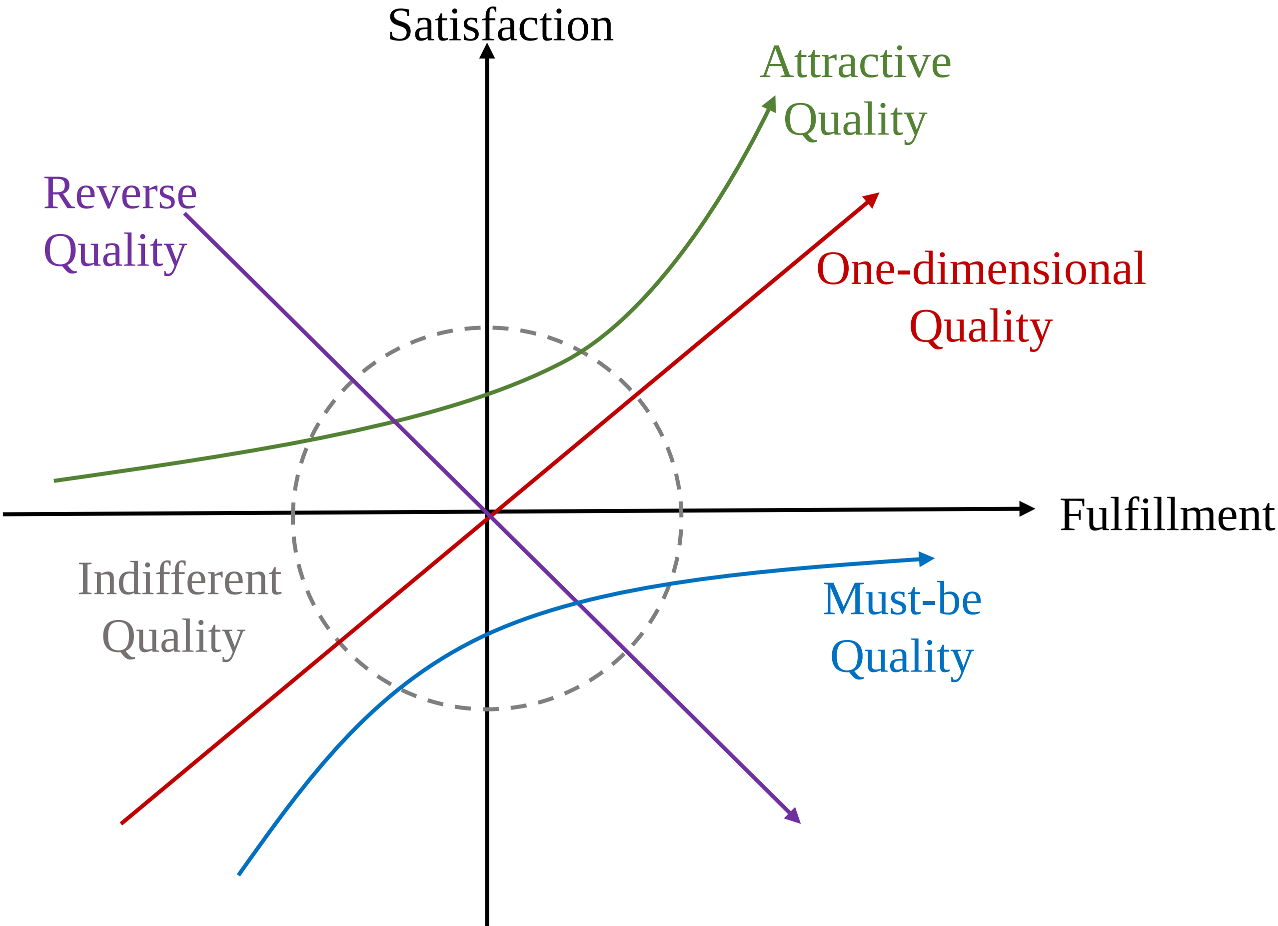
- Original Dataset: 500 cruise reviews collected between February 2019 and December 2024, which can be found in the supplementary file labeled "raw_cruise_reviews_copy_20241121".
- Processed Data: The corresponding sentiment analysis output data is available in the supplementary file labeled "SA_output_data".

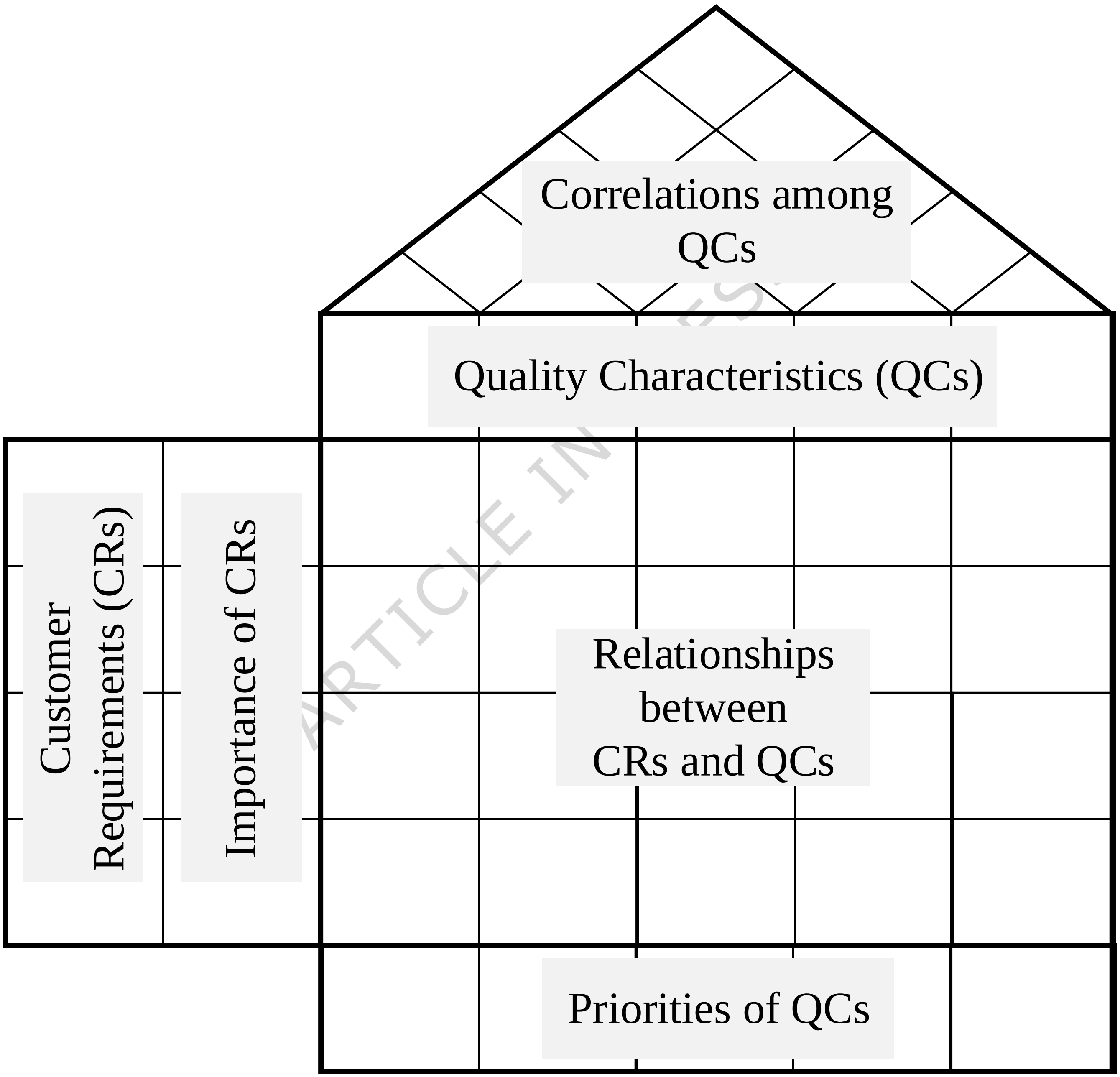
Ethical Approval

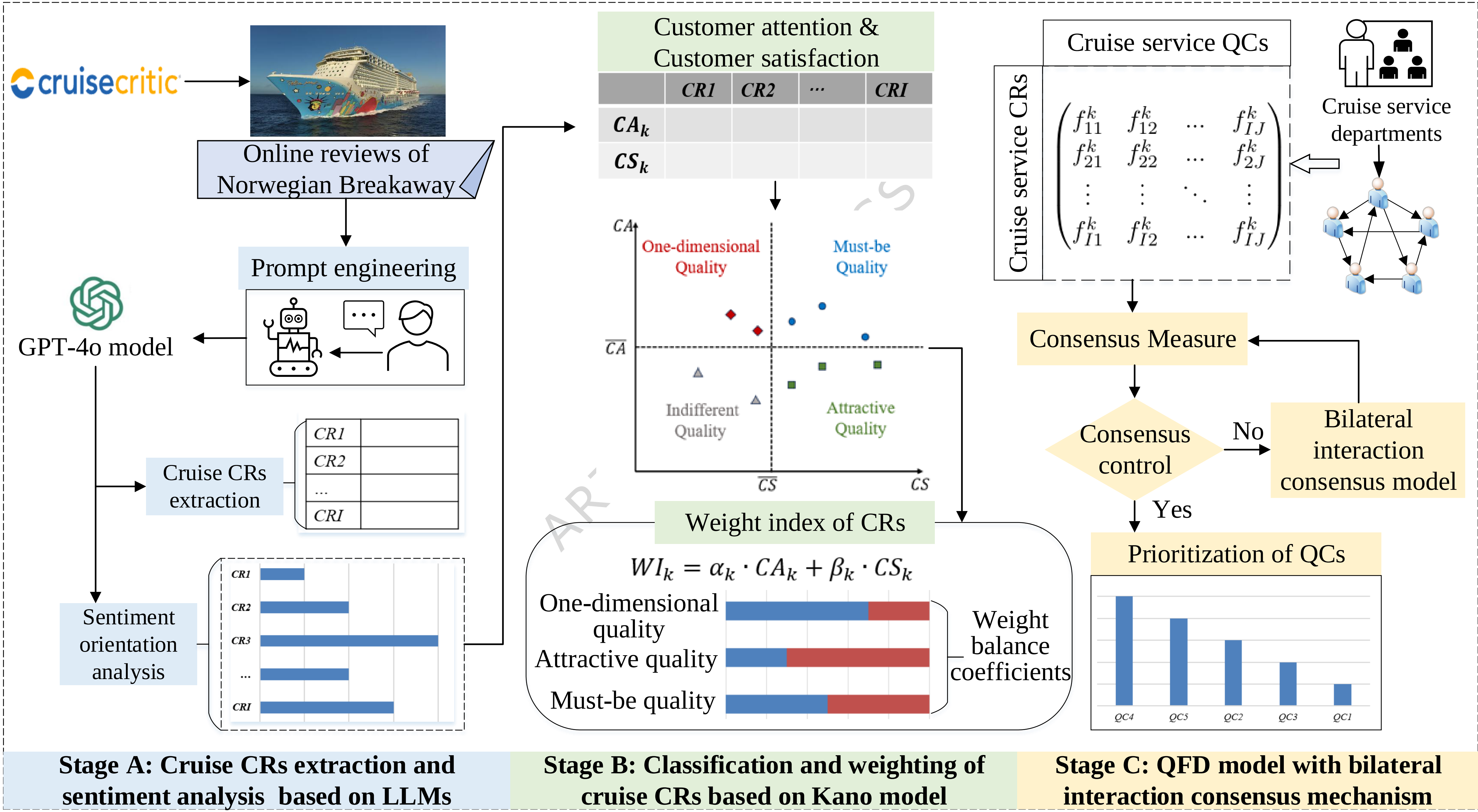
Ethical approval was not required for this study. This research does not involve human participants, experimental interventions, or personally identifiable data. The study relies exclusively on the observational, secondary analysis of 500 publicly available cruise reviews crawled from a public website between February 2019 and December 2024. According to the Trial Measures for Ethical Review of Science and Technology (issued by the Ministry of Science and Technology of the People's Republic of China in 2023), research involving the secondary analysis of existing, legally obtained, and publicly accessible data where human subjects cannot be identified is not classified as human subjects research requiring ethical oversight. Because this study falls outside the regulatory scope of institutional ethical review, formal ethical approval and institutional waiver processes are not applicable.

Informed Consent

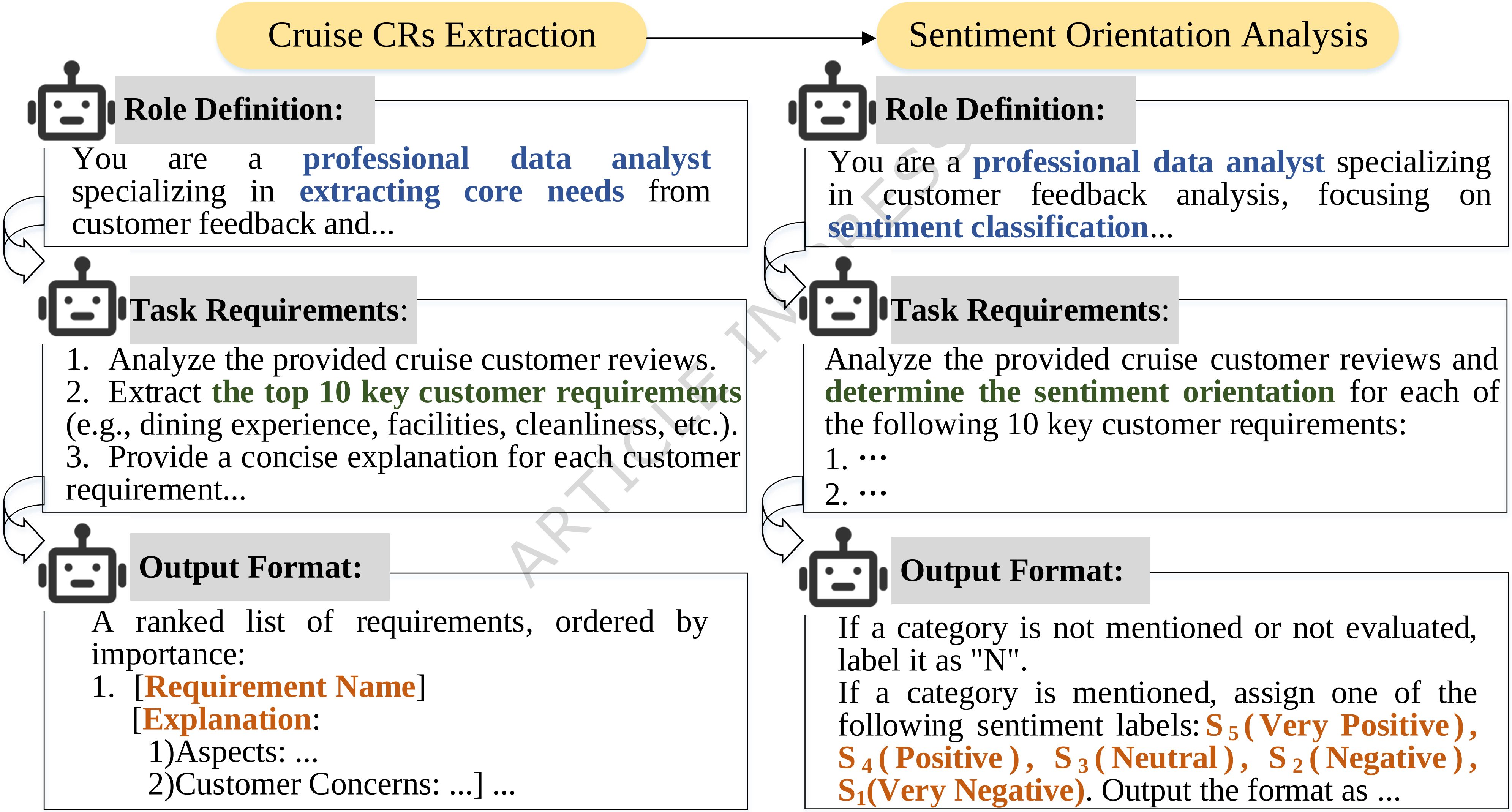
Informed consent was not applicable for this study. The research does not involve direct interaction with any individuals. The dataset consists solely of pre-existing, publicly accessible online reviews voluntarily posted by users in the public domain. According to Article 27 of the Personal Information Protection Law of the People's Republic of China (PIPL), personal information handlers may process personal information disclosed by the individuals themselves or otherwise legally disclosed within a reasonable scope without requiring explicit consent. All data extracted between February 2019 and December 2024 were strictly anonymized prior to analysis, and no personally identifiable information (PII) was collected. Therefore, the requirement for obtaining informed consent is legally and practically waived.

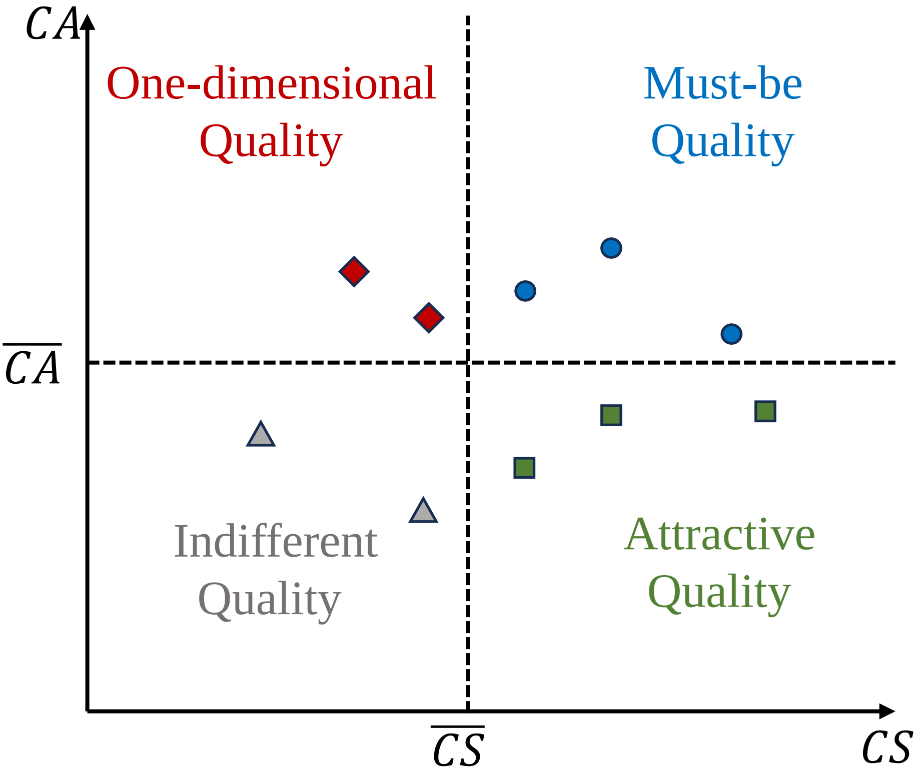

Satisfaction
Attractive
Quality
Reverse
Quality
One-dimensional
Quality
Fulfillment
Indifferent
Quality
Must-be
Quality


Correlations among QCs
Quality Characteristics (QCs)
Customer Requirements (CRs)
Importance of CRs
Relationships between CRs and QCs
Priorities of QCs

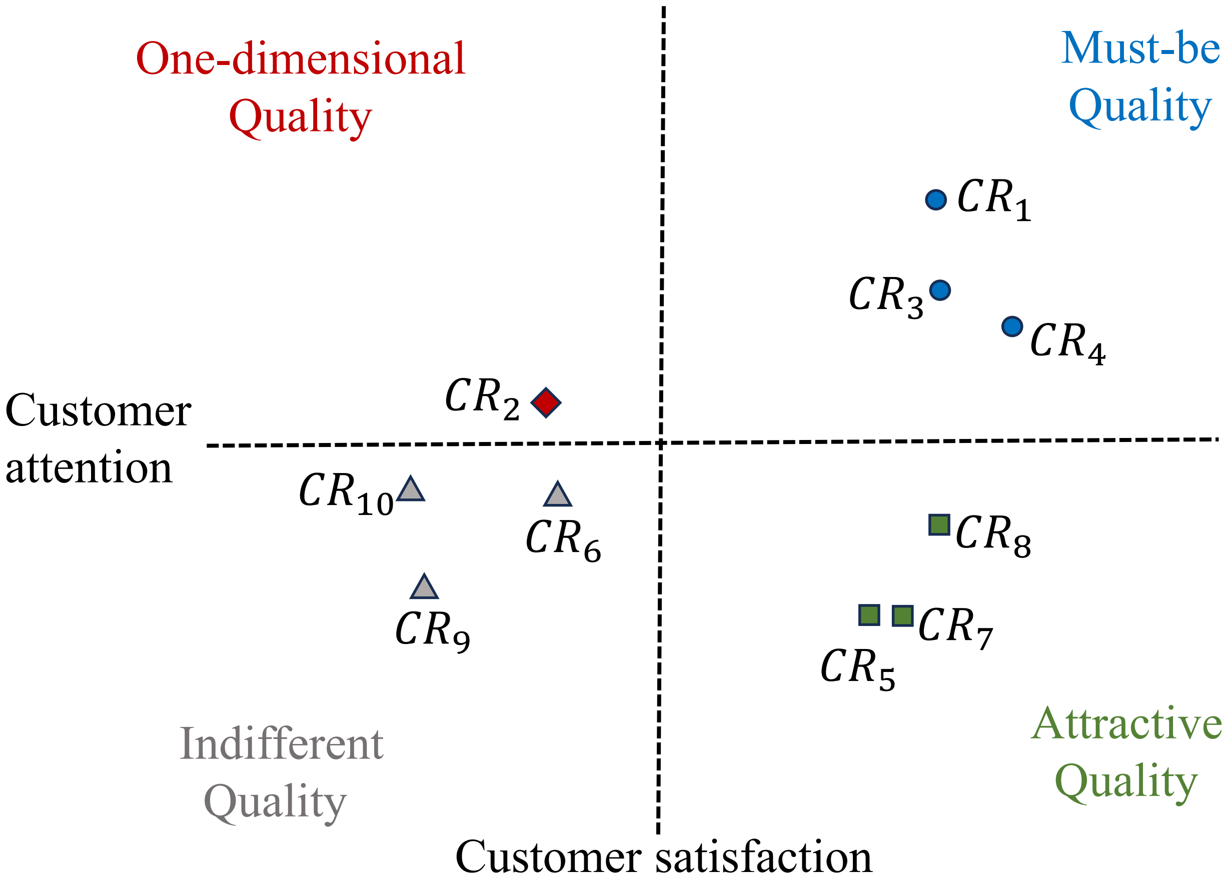


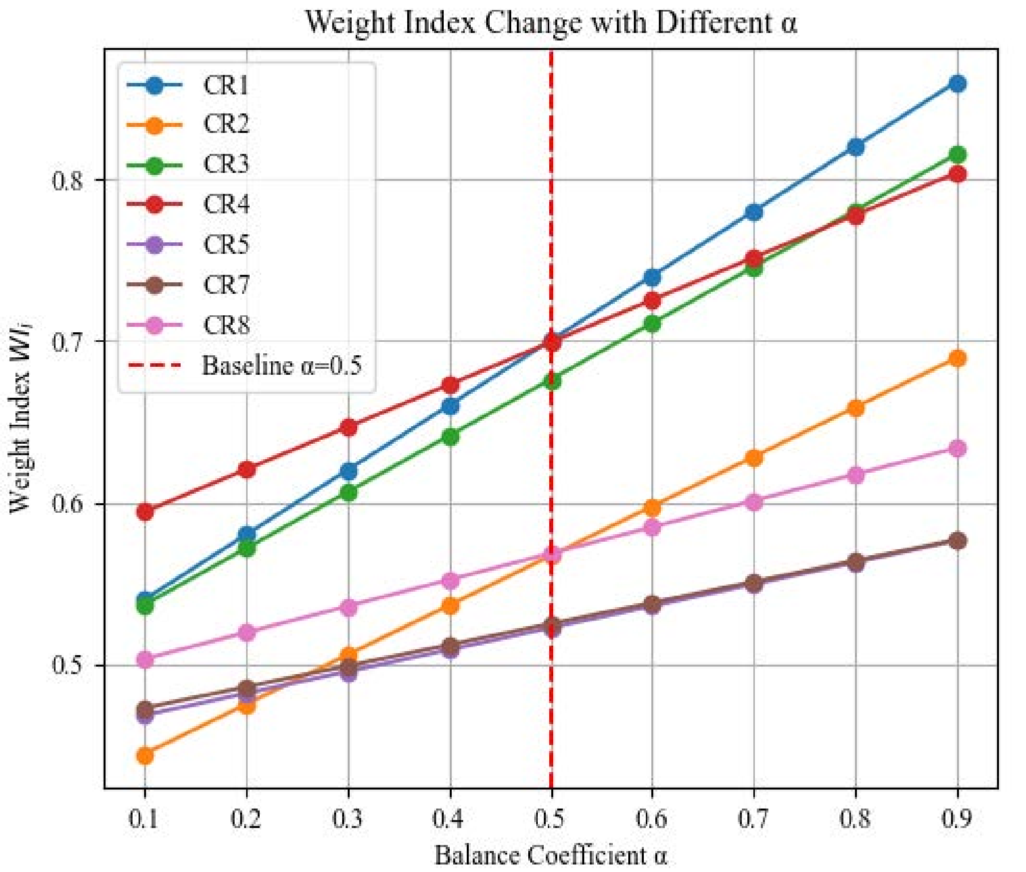
cruisecritic
Online reviews of Norwegian Breakaway
Prompt engineering
GPT-4o model
Cruise CRs extraction
CR1
CR2
...
CRI
Sentiment orientation analysis
CR1
CR2
CR3
...
CRI
Customer attention & Customer satisfaction
CR1
CR2
...
CRI
CA_k
CS_k
CA
One-dimensional Quality
Must-be Quality
$\overline{CA}$
Indifferent Quality
Attractive Quality
$\overline{CS}$
CS
Weight index of CRs
$WI_k = \alpha_k \cdot CA_k + \beta_k \cdot CS_k$
One-dimensional quality
Attractive quality
Must-be quality
Weight balance coefficients
Cruise service QCs
Cruise service CRs
$\begin{pmatrix} f_{11}^k & f_{12}^k & \cdots & f_{1J}^k \\ f_{21}^k & f_{22}^k & \cdots & f_{2J}^k \\ \vdots & \vdots & \ddots & \vdots \\ f_{I1}^k & f_{I2}^k & \cdots & f_{IJ}^k \end{pmatrix}$
Cruise service departments
Consensus Measure
Consensus control
No
Bilateral interaction consensus model
Yes
Prioritization of QCs
QC4
QC5
QC2
QC3
QC1
Stage A: Cruise CRs extraction and sentiment analysis based on LLMs
Stage B: Classification and weighting of cruise CRs based on Kano model
Stage C: QFD model with bilateral interaction consensus mechanism

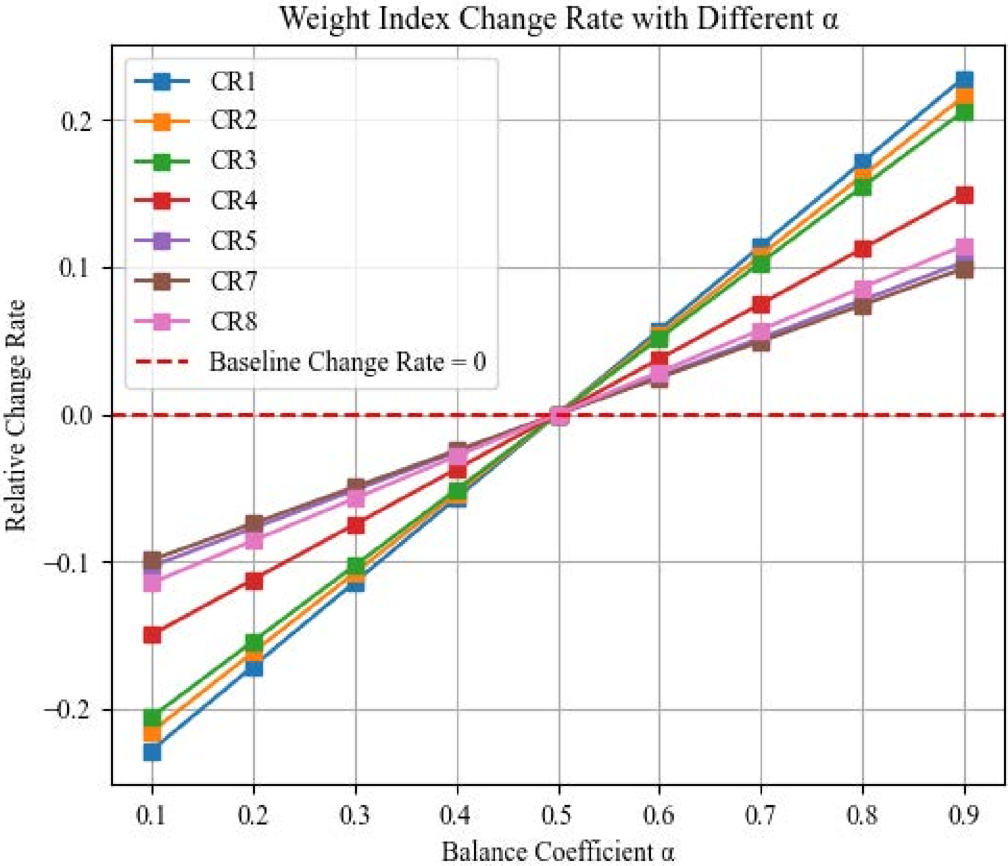

Cruise CRs Extraction
Sentiment Orientation Analysis
Role Definition:
You are a professional data analyst specializing in extracting core needs from customer feedback and...
Task Requirements:
1. Analyze the provided cruise customer reviews.
2. Extract the top 10 key customer requirements (e.g., dining experience, facilities, cleanliness, etc.).
3. Provide a concise explanation for each customer requirement...
Output Format:
A ranked list of requirements, ordered by importance:
1. [Requirement Name]
[Explanation:
1)Aspects: ...
2)Customer Concerns: ...] ...
Role Definition:
You are a professional data analyst specializing in customer feedback analysis, focusing on sentiment classification...
Task Requirements:
Analyze the provided cruise customer reviews and determine the sentiment orientation for each of the following 10 key customer requirements:
1. ···
2. ···
Output Format:
If a category is not mentioned or not evaluated, label it as "N".
If a category is mentioned, assign one of the following sentiment labels: S5 (Very Positive), S4 (Positive), S3 (Neutral), S2 (Negative), S1(Very Negative). Output the format as ...

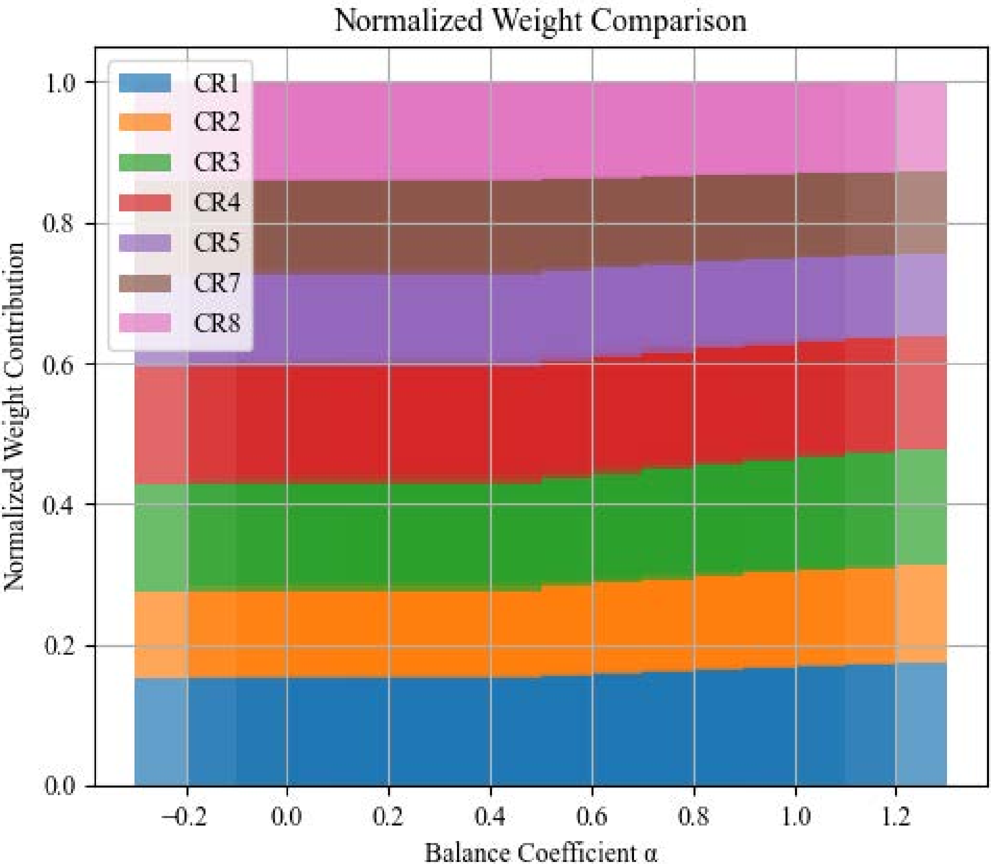


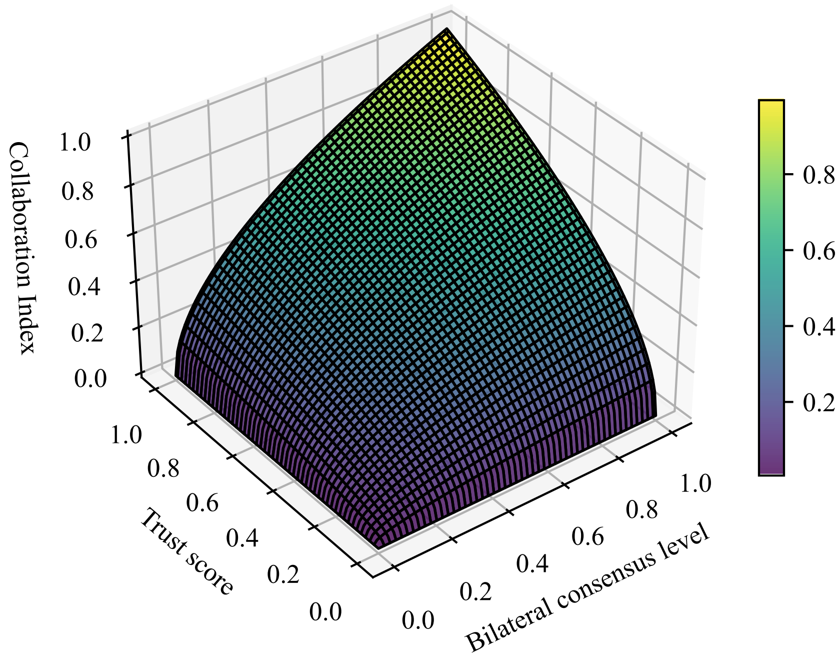
CA
One-dimensional Quality
Must-be Quality
$\overline{CA}$
Indifferent Quality
Attractive Quality
$\overline{CS}$
CS

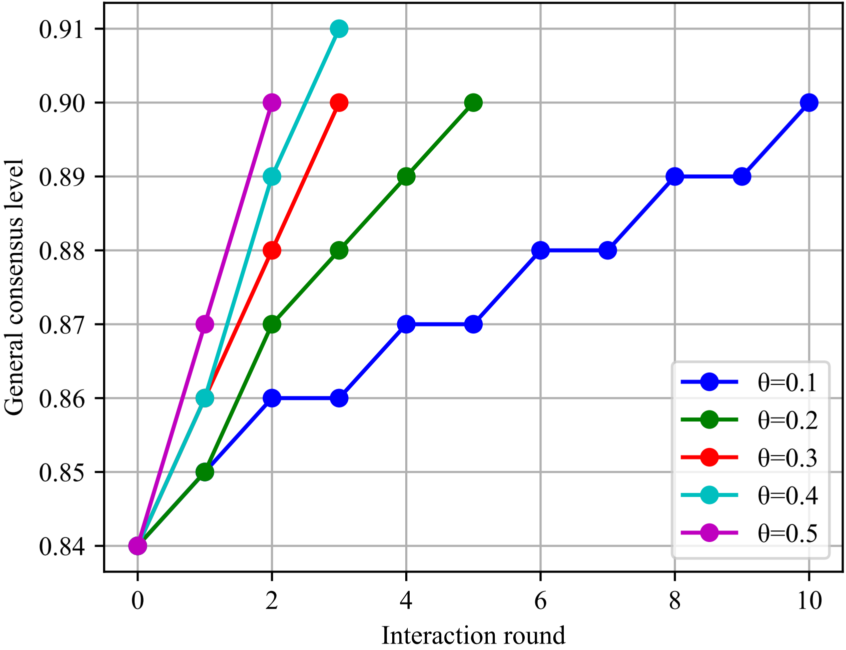

One-dimensional Quality
Must-be Quality
CR_1
CR_3
CR_4
CR_2
Customer attention
CR_{10}
CR_6
CR_8
CR_9
CR_7
CR_5
Indifferent Quality
Attractive Quality
Customer satisfaction


Weight Index Change with Different α
CR1
CR2
CR3
CR4
CR5
CR7
CR8
Baseline α=0.5
Weight Index W_i
0.5
0.6
0.7
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Balance Coefficient α


Weight Index Change Rate with Different α
CR1
CR2
CR3
CR4
CR5
CR7
CR8
Baseline Change Rate = 0
Relative Change Rate
0.2
0.1
0.0
−0.1
−0.2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Balance Coefficient α


Normalized Weight Comparison
CR1
CR2
CR3
CR4
CR5
CR7
CR8
Normalized Weight Contribution
0.0
0.2
0.4
0.6
0.8
1.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Balance Coefficient α


Collaboration Index
1.0
0.8
0.6
0.4
0.2
0.0
Trust score
Bilateral consensus level


0.91
0.90
0.89
0.88
0.87
0.86
0.85
0.84
General consensus level
0
2
4
6
8
10
Interaction round
θ=0.1
θ=0.2
θ=0.3
θ=0.4
θ=0.5